SAPIENZA
UNIVERSITÀ DI ROMA

DOTTORATO DI RICERCA IN STATISTICA ECONOMICA

XXV CICLO

Essays on Econometrics of Stategic Interactions

Supervisor:

Eleonora Patacchini

Candidate:

Edoardo Rainone

A.A. 2014/2015

Acknowledgements

I would like to thank all the people that helped and supported me with writing this dissertation. First, I want to thank Eleonora Patacchini for her great support and encouragement during these years, for driving me through the topics and for continuously stimulating me with her unlimited energy.

Many thanks go to my friend and co-author Tiziano Arduini for all his precious work and all the inspiring discussions about economic ideas and econometric problems.

Further, I am particularly grateful to my co-authors Xiaodong Liu and Yves Zenou, for all the expertises and knowledge they transmitted to me.

I would like to thank Carlo Leone del Bello, Pallab Ghosh, Lincoln H. Groves, Nuno Mota, Judson Murchie, Sun Jung Oh, Jan Ondrich, Ian Wright, Pengju Zhang for all the constructive conversations at Syracuse University during my visiting period, and Desislava Miroslavova Dankova, Alessia Fulvimari, Gianluigi Nocella, Eleonora Romano for all the good study time spent together at Sapienza Universita' di Roma.

I thank Salvatore Luigi Alonzo and Fabrizio Palmisani for giving me the opportunity to visit Syracuse University and my colleagues at Banca d'Italia for the endurance they had, being continuously bothered by an annoying PhD Student.

I have to thank Enrica Siracusa for being so patient with a partner often spending his leisure time writing his thesis, for supporting me unconditionally and for loving me so much.

My final and most important thank goes to my parents, Maria Luisa Marzetti and Antonio Rainone, who taught me to be curious and enthusiastic about things, always preserving patience and dedication.

All the ingredients needed to do research.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This dissertation is focused on the econometrics of strategic interactions. When an economic decision has to be analyzed, agents can be modeled as isolated decision makers. It eases models' tractability at the cost of being not so close to reality. A growing literature on economic theory and econometrics is concerned with this aspect (see Jackson and Zenou; 2014; Benhabib et al.; 2011 for a review). The dissertation is aimed to contribute to this branch of literature, in particular to theoretical and empirical econometrics.

Interactions can be embedded in financial and economic models in several ways. In a model with strategic interactions the outcome of each agent usually depends on a function of other agents' outcomes, the so-called *endogenous effect*, and on other agents' characteristics, the *contextual effect*. An example is the famous linear-in-mean model, in which the outcome of each agent depends linearly on the average outcome of the group and the average characteristics of the same group. Several issues arise with the estimation of interaction models. In his seminal work Manski (1993) outlined the so-called reflection problem for the linear-in-mean model. It can be overcome if interactions are sparse, i.e. if they can be represented by a incomplete network, as Bramoullé et al. (2009) have shown. In other words, in such a *network model* the reference group is not the

same for each agent in the group,[1] this creates non-linearities which in turn allow to identify model's parameters. The econometrics behind this class of models has the same foundations of the spatial econometric literature see e.g. Lee et al. (2010).

In this dissertation I will consider interaction models with network structures. In particular, the specifications considered in this study belong to the family of spatial autoregressive (SAR) models, in which the information on networks is summarized in a adjacency matrix, see Cliff and Ord (1973) for a survey.

Even though the reflection problem is not at work for this family of models, other issues are related to the estimation of parameters.

This work focuses on the three of those: *heterogeneity* of the endogeneous effects, *endogeneity* of the interactions, and *unobservability* of interacting agents. Each extension is discussed in a chapter of my dissertation. The first chapter is a purely theoretical one, the last two also contain empirical applications with a focus on financial and economic decisions.

The first chapter is about the possible heterogeneity of endogeneous effects. It generalizes a network model where different groups of individuals are allowed to exert a different influence and where both within and between groups interactions can be at work. In particular it studies the identification and estimation of treatment response with heterogeneous spillovers in a network model. It generalizes the standard linear-in-means model to allow for multiple groups with between and within-group interactions. It provides a set of identification conditions of peer effects and consider a 2SLS estimation approach. Large sample properties of the proposed estimators are derived. Simulation experiments show that the estimators perform well in finite samples. The model is used to study the effectiveness of policies where peer effects are seen as a mechanism through which the treatments could propagate through the network. When interactions among

[1]Note that the linear-in-mean model is a special case of a network model, when the network is full, i.e. everybody is connected with everybody.

groups are at work, a shock on a treated group has effects on the non-treated. Its framework allows for quantifying how much of the indirect treatment effect is due to variations in the characteristics of treated peers (treatment contextual effects) and how much is because of variations in peer outcomes (peer effects).

The second chapter studies the importance of social interactions for the adoption of financial products using a bayesian methodology in order to account for network endogeneity. It exploits a unique dataset of friendships among United States students and a novel estimation strategy that accounts for possibly endogenous network formation. It finds that not all social contacts are equally important: only those with a long-lasting relationship influence financial decisions. Moreover, the correlation in agents' behavior only arises among long-lasting ties in cohesive network structures. This evidence is consistent with an important role of trust in financial decisions. Repeated interactions generate trust among agents, which in turn aggregate in tightly knit groups. When agents have to decide whether or not to adopt a financial instrument they face a risk and might place greater value on information coming from agents they trust. These results can help to understand the growing importance of face-to-face social contacts for financial decisions.

The third chapter is concerned with the consistent estimation of parameters in a network model when the outcome is not observed for every agent belonging to the network. An application on the allocation of time in sleep by agents is also provided. Sleep is a key determinant of educational attainment among young adults, and carries with it longstanding health implications. It provides evidence of network effects in adolescent sleeping decisions and develop a novel econometric approach to estimate network models with sampled observations on the dependent variable. When accounting for sampling, it finds that the sleeping behaviour of the friends is important to shape own sleeping behaviour, besides the impact of individual, family and friends characteristics. Unique information

on siblings and their friends allows us to check the robustness of our results to unobserved family factors.

Chapter 2

Identification and Estimation of Outcome Response with Heterogeneous Treatment Externalities

Joint work with Tiziano Arduini and Eleonora Patacchini.

Another version of this chapter has been published as *CPR Working Paper No. 167, April 2014.*

2.1 Introduction

The program evaluation literature focuses on estimating the program effects without externalities. There is a growing awareness, however, that there may be indirect effects that are important to measure (see Manski (2013)). Existing methodological contributions as well as studies collecting empirical evidence are still scarce. In particular, while there are a few papers about the identification and estimation of treatment response with interactions (Hudgens and Halloran

(2008), Miguel and Kremer (2004) and Sinclair et al. (2012)), to the best of our knowledge there are no studies that consider the presence of *heterogeneous interactions.*

Angelucci and De Giorgi (2009) estimates the indirect effects of the flagship Mexican welfare program, PROGRESA, on the consumption of ineligible households. This study finds that cash transfers to eligible households indirectly increase the consumption of ineligible households living in the same village. These findings are clearly very important for designing policies as well as developing experiments to evaluate them.[1] The framework, however, does not determine how much of the spillover is due to effects from eligible to ineligible subjects, effects within ineligible (eligible) subjects and feedback effects. It identifies the presence of indirect effects by comparing outcomes between untreated household in untreated villages and untreated households in treated villages. When network data are available, the analysis can be pushed forward and the heterogeneous impact of policies can be modeled and quantified.

Heterogeneity can be conceived in different ways. First, *treatment heterogeneity*, when the intensity or type of treatment can differ depending on the treated unit. Second, *treatment effect heterogeneity* when the treatment is the same for each agent but its effect is different depending on her characteristics. Third, *interaction-driven heterogeneity*, when the diffusion of the treatment effect through interactions generates an heterogeneous individual response. This may be due to both differences in interaction strengths within and between groups and to network structure, if data on connections are available. Several papers have focused on the first two types of heterogeneity.[2] In this paper we focus on

[1] More specifically, policy interventions should internalize the externalities that they engender, and experiments to evaluate their effectiveness should consider the effects on the entire local economy (e.g. the school, the village, the city), rather than focusing on differences between treatment and control group from the same local entity. When spillovers are at work, both groups' performance may change.

[2] See Imbens and Woolridge (2009) for a revision of recent studies using matching and non-parametric methodologies to address the second type of heterogeneity. Remarkably, Crump et al. (2008) proposes a non-parametric test for subpopulation heterogeneity in the effect of the

the third kind of heterogeneity.

Using a network approach, our analysis brings three contributions to this literature. First, we derive analytically the bias that arises if spillovers are ignored. Second, we provide estimands for understanding whether different types of untreated - eligible or ineligible- are differently impacted by the treatment. Finally, our framework allows us to distinguish between different sources of treatment transmission - in particular, how much of the treatment response is generated by variations in the charactristics of treated peers(*treatment contextual effects*) and how much is due to spillovers through outcomes (*peer effects*). More specifically, our paper provides a network-based approach to estimate the average effects of the treatment in the presence of spillovers on subjects both eligible and ineligible for a program, accounting for heterogeneous within and between-group spillover effects. We show that heterogeneity in the effects is both helpful in terms of identification and harmful for traditional estimation methods. We develop an estimation approach able to provide reliable estimates of all the cascade effects that stem from a given network topology.

Interaction among agents can be modeled in several ways. When the exact topology of connections is know, one possibility is to consider the peer effects that stem from the given network structure. There is a large and growing literature on peer effects in economics using network data.[3] The popular model employed in empirical work is the Manski-type linear-in-means model (Manski (1993)). Three assumptions underlie this statistical model: *(i)* the network is exogenous, *(ii)* the effects of all peers are equal, *(iii)* peer status is measured without error. Although these assumptions may be restrictive in empirical anal-

treatment. Firpo (2007) proposed a quantile treatment estimation where the heterogeneity is given by the position of unit in the pre-treatment outcome distribution. Other papers employ more complex techniques to allow both the first and second type of heterogeneity. Among the others, generalized cross-validation statistic (Imai et al. (2013)), boosting (LeBlanc and Kooperberg (2010)), Bayesian Additive Regression Trees (Chipman et al. (2010)) have been used.

[3]See Jackson and Zenou (2014), part III, for a collection of recent studies.

yses, only a few recent papers consider alternative models and methods in which some of these assumptions are relaxed. Point *(i)* has been recently studied by Goldsmith-Pinkham and Imbens (2013) and Hsieh and Lee (2011) who propose parametric modelling assumptions and Bayesian inferential methods to integrate a network formation model with the study of behavior over the formed networks. Point *(iii)* belongs to another strand of recent literature which looks at the consequences of peer-group misspecification, focusing in particular on sampling issues (see Chandrasekhar and Lewis (2011a), Liu et al. (2013) and Liu (2013a)). In this paper, we consider the specification and estimation of a peer effects model when assumption *(ii)* is removed. Lee and Liu (2010) considers a peer effects model with one endogenous variable and one adjacency matrix in a multiple network context, with no between-network interactions. Liu (2013) extends this model to the case of two endogenous variables and one adjacency matrix. In this paper, we allow the model to have two endogenous variables, two adjacency matrices, and both within and between-group interactions. We also consider the generalization to the case of multiple endogenous variables.[4] To the best of our knowledge, we are the first to consider models of peer effects where different peers are allowed to exert a different influence and where both within and between groups interactions can be at work.[5] We maintain assumptions *(i)* and *(iii)*.

We show that the multiple group structure of the model requires modifying the conventional identification conditions (Bramoullé et al. (2009) and Cohen-Cole et al. (2012)) and has interesting connections with the concepts of *chains* and *Tree-indexed Markov chains* (see Benjamini and Peres (1994)).

We propose efficient 2SLS estimators using instruments based on the two reduced forms. We show that the standard IV approximation (Kelejian and

[4]There is a long tradition in spatial econometrics looking at spatial autoregressive models with multiple endogenous variables (see Kelejan and Prucha, 2004). In the spatial econometrics context, however, the adjacency matrix is the same for all endogenous variables, and no groups are considered.

[5]Goldsmith-Pinkham and Imbens (2013) also estimate a model with two peer effects, but without cross effects, using a Bayesian estimation method.

Prucha (1998), Kelejian and Prucha (1999) and Liu and Lee (2010)) involves a huge number of IVs, even if we use a low degree approximation of the optimal instruments.[6] For this reason, we consider many-instrument asymptotics (Bekker, 1994) allowing the number of IVs to increase with the sample size.

Differently from Lee and Liu (2010) and Liu (2013b) where the many instruments derive from the multiple network framework, in our model the many instruments derive from the (approximation of the) multiple adjacency-matrix framework. A multiple matrix framework does not only result in an increasing number of instruments but also yields multiple approximations of the optimal instruments. As a result, we show that the form of the many-instrument bias differs, though the leading order remains unchanged. We also propose a bias-correction procedure. Simulation experiments show that the bias-corrected estimator performs well in finite samples. When the number of endogenous variables is allowed to grow, our estimator remains consistent and asymptotically normal if the number of endogenous variables grows more slowly than the sample size. Finally, we investigate the bias occurring when the interaction structure is misspecified. We derive analytically the bias that occurs when only within-group peer effects are considered, i.e when interactions between groups are at work but ignored by the econometrician. We then use a simulation experiment to evaluate this bias in finite samples.

In the last part of the paper we show the empirical salience of our model for policy purposes. As highlighted by Manski (2013), the policy maker can rarely manipulate peer outcomes. Peer effects, however, can be seen as a mechanism through which the treatment could propagate through the networks. If peer effects are at work, then the policy intervention has not only a direct effect on outcomes but also an indirect one through the outcomes of connected agents (i.e. the so called "social multiplier"). We show via Monte Carlo simulations that

[6] See Prucha (2013) for a review of Generalized Method of Moments estimators in a spatial framework.

the presence of *heterogeneous peer effects* and *between-group interactions* may create unexpected, or sometimes paradoxical results if the policy maker ignores the heterogeneity of interactions among groups. Our results can be helpful to explain why several policy programs do not accomplish the expected goals.

The paper is organized as follows. The next section introduces the econometric model. Identification conditions are derived in Section 3, and in Section 4 we consider 2SLS estimation for the model. Section 5 investigates the bias occurring when the interaction structure is misspecified. We devote Section 6 to show the importance of our analysis for the identification of treatment response with spillovers. We first derive estimands for direct, indirect and total effects of treatment strategies in network settings with interactions. Then we use a simulation experiment to show the extent to which the heterogeneity of the endogenous effects can affect the outcome response for different groups. Section 7 concludes.

2.2 The Network Model with Heterogeneous Peer Effects

A general network model has the specification

$$Y = \phi GY + X\beta + G^*X\gamma + \epsilon, \tag{2.1}$$

where $Y = (y_1, \ldots, y_n)'$ is an $n-$dimensional vector of outcomes, $G = [g_{ij}]$ is an $n \times n$ adjacency matrix, g_{ij} is equal to 1 if i and j are connected, 0 otherwise. G^* is the row-normalized version of G, where $g^*_{ij} = g_{ij}/\sum_j g_{ij}$. X is a $n \times p$ matrix of exogenous variables capturing individual characteristics. $\epsilon = (\epsilon_1, \ldots, \epsilon_n)'$ is a vector of errors whose elements are i.i.d. with zero mean and variance σ^2 for all i. For model (2.1), ϕ represents *the endogenous effect*, where an agent's choice/outcome may depend on those of his/her peers on the same activity, and

γ represents *the contextual effect*, where an agent's choice/outcome may depend on the exogenous characteristics of his/her peers. Let $X^* = (X, G^*X)$ and $\beta^* = (\beta, \gamma)$.

Let A and B be two countable sets (types) of individuals (e.g. males and females, blacks and whites) such that $A \cap B = \varnothing$ and $n = n_a + n_b$ is the cardinality of $A \cup B$, with n_a and n_b being respectively the cardinalities of A and B. Let us define $Y = (Y_a', Y_b')'$, $X = (X_a', X_b')'$, and $G = \begin{bmatrix} G_a & G_{ab} \\ G_{ba} & G_b \end{bmatrix}$. For instance, the subscript a denotes that $Y, X \in A$, G is formed only among nodes of type A and the subscript ab denotes the fact that links are directed from b to a.[7] Appendix A defines regularity conditions.

Model (2.1) can be written as

$$Y_a = \phi_a G_a Y_a + \phi_{ab} G_{ab} Y_b + X_a^* \beta_a^* + G_{ab}^* X_b \gamma_{ab} + \epsilon_a, \tag{2.2}$$

$$Y_b = \phi_b G_b Y_b + \phi_{ba} G_{ba} Y_a + X_b^* \beta_b^* + G_{ba}^* X_a \gamma_{ba} + \epsilon_b, \tag{2.3}$$

where $\beta_a^* = (\beta_a, \gamma_a)$, $X_a^* = (X_a, G_a^* X_a)$, $X_b^* = (X_b, G_b^* X_b)$ $\beta_b^* = (\beta_b, \gamma_b)$, and ϵ_a and ϵ_b are i.i.d errors with variance σ_a^2 and σ_b^2, respectively. Let us suppose for simplicity that $\sigma_a^2 = \sigma_b^2 = \sigma$. Model (2.2) - (2.3) is a generalization of the standard framework in the sense that it allows endogenous effects to be different within and between groups. If we stack up equations (2.2) - (2.3) and restrict the endogenous effect parameters of the two equations to be the same (i.e. $\phi_a = \phi_b = \phi_{ab} = \phi_{ba}$), then we obtain model (2.1).

Let us define the following matrices

$$A\delta_a = X_a^* \beta_a^* + G_{ab}^* X_b \gamma_{ab} + \epsilon_a,$$

[7] More formally, $Y_a = R_a Y$, $X_a = R_a X$, $G_a = R_a G R_a'$ and $G_{ab} = R_a G R_b'$, where $R_a = (I_{n_a}, O_{n_a, n_b})$ and $R_b = (O_{n_b, n_a}, I_{n_b})$ are matrices that select the nodes in group a and b respectively. $O_{k,l}$ is a $k \times l$ matrix of zeros.

$$B\delta_b = X_b^*\beta_b^* + G_{ba}^* X_a \gamma_{ba} + \epsilon_b,$$

where $A = (X_a, G_{ab}^* X_b, \epsilon_a)$, $\delta_a = (\beta_a^*, \gamma_{ab}, 1)$, $B = (X_b, G_{ba}^* X_a, \epsilon_b)$ and $\delta_b = (\beta_b^*, \gamma_{ba}, 1)$. By plugging Y_b in equation (2.2) we have

$$\begin{aligned} Y_a &= \phi_a G_a Y_a + \phi_{ab} G_{ab}(J_b(\phi_{ba} G_{ba} Y_a + B\delta_b)) + A\delta_a \\ &= (\phi_a G_a + \phi_{ab}\phi_{ba} C_a) Y_a + \phi_{ab} G_{ab} J_b B\delta_b + A\delta_a, \end{aligned} \tag{2.4}$$

where $J_b = (I - \phi_b G_b)^{-1} = \sum_{k=1}^{\infty} (\phi_b G_b)^k$ provided $\|\phi_b G_b\|_\infty < 1$, where $\|\cdot\|_\infty$ is the row-sum matrix norm. The ij_{th} element of J_b sums all k-distance paths from j to i when $i, j \in B$ scaling them by ϕ_b^k and $C_a = G_{ab} J_b G_{ba}$.[8] Therefore the reduced form of model (2.2) is

$$Y_a = M_a(\phi_{ab} G_{ab} J_b B\delta_b + A\delta_a), \tag{2.5}$$

where $M_a = (I - \phi_a G_a - \phi_{ab}\phi_{ba} C_a)^{-1}$.[9] A sufficient condition for the non singularity of $(I - \phi_a G_a - \phi_{ab}\phi_{ba} C_a)$ is $\|\phi_a G_a\|_\infty + \|\phi_{ab}\phi_{ba} C_a\|_\infty \leq 1$. This condition also implies that M_a is uniformly bounded in absolute value.[10]

We note that: *(i)* we present an aggregate model specification (i.e. G which multiplies y in model (2.1) is not row-normalized), but the approach applies also to an average model (i.e. when G which multiplies y in model (2.1) is row-normalized);[11] *(ii)* our model specification has two groups, but all the assump-

[8] C_a is a matrix which captures all the indirect connections among nodes of type A passing through one or more nodes of type B. Note that the ij_{th} generic element of $G_{ab}G_{ba}$ is equal to the number of length-2 paths directed from $j \in A$ to $i \in A$ passing through a node $l \in B$. This matrix accounts only for distance-2 indirect connections while $C_a = G_{ab} J_b G_{ba}$ captures all the paths starting from $j \in A$ and ending to a generic node in B, eventually passing through other nodes of type B and finally arriving in $i \in A$ scaling them by ϕ_b.

[9] This matrix captures all direct and indirect paths among type A nodes passing through others type A nodes and type B nodes.

[10] The assumption is crucial for identification of the model and asymptotic normality of the estimator (see Appendix A).

[11] Aggregate and average models are different in terms of behavioral foundations, contextual effects are supposed to be averages over peers in both cases w.l.o.g. (see Liu et al. (2014forth-

tions, propositions and proofs can be naturally extended to a finite number of groups; *(iii)* we consider a single network, but the approach can be extended to the case of multiple networks(i.e. a network with several components) with the addition of network fixed effects in the model specification; *(iv)* we can also add a heterogeneous spatial lag in the error term $\epsilon_a = \rho_a W_a \epsilon_a + \rho_{ab} W_{ab} \epsilon_b$.[12]

2.3 Identification

Let us define $Z_a = (G_a Y_a, G_{ab} Y_b, X_a^*, G_{ab}^* X_b)$. Equation (2.2) is identified if $E(Z_a)$ has full column rank for large n.[13] In this section, we find sufficient conditions for $E(Z_a)$ to have full column rank.[14] The detailed proof is given in Appendix C.

Proposition 1. Let X_a and X_b have full column rank. If the sequences of $\{M_a\}$, $\{M_b\}$, $\{J_a\}$ and $\{J_a\}$ are UB matrices,[15] then $E(Z_a)$ has full column rank in the following cases

1. (a) i. $\beta_a \phi_a + \gamma_a \neq 0$,

   ii. I_a, G_a and G_a^2 are linearly independent.

   [and]

   (b) i. $\beta_b \phi_b + \gamma_b \neq 0$,

   ii. G_{ab} and $G_{ab} G_b$ are linearly independent.

   [or]

2. (a) i. $\gamma_{ab} \neq 0$,

   ii. G_{ab} and $G_a G_{ab}$ are linearly independent.

coming)).

[12] The resulting model is a $SARARMAG(p; q; g)$ with $p = 1$, $q = 1$ and $g = 2$, where p and q are respectively the number of spatial lags for outcome and error, and g is the number of groups (see Kelejian and Prucha (2007)).

[13] This implies that Assumption 4 in Appendix A holds.

[14] Symmetric conditions and results hold for equation (2.3).

[15] In practice we need a series expansion to approximate the inversion of the matrices. We are grateful to Chihwa Kao for pointing it out.

[and]

(b) i. $\gamma_{ba} \neq 0$,

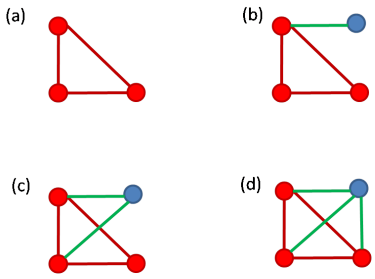
ii. I_a, G_a and $G_{ab}G_{ba}$ are linearly independent.

Note that conditions (1a) are exactly the same identification conditions found by Bramoullé et al. (2009) in the case of homogeneous effects (i.e. only one group). Proposition 1 here is more general as it provides alternative possibilities. When more than one group is considered we do not need linear independence of a particular set of matrices - we have multiple sufficient conditions. Even if I_a, G_a and G_a^2 are linearly dependent we can still identify ϕ_a, and the other parameters, relying on linear independence of chains passing through type B nodes.[16] The set of adjacency matrices' combinations can be represented as a *Tree-indexed Markov chain*- the parameters can be identified because of the multiple branches of the tree (see Appendix B). Obviously, if G_a, G_{ba}, G_{ab} and G_b are complete and consequently all products among them are linearly dependent, then the model remains not identified. However, if group A nodes are in a complete network, but the matrices representing between-group interactions are sparse (i.e G_{ab} and G_{ba} are not complete), then identification can be achieved and ϕ_a can be estimated even if G_a is complete. Systems in panels (b) and (c) in Figure 2.1 can be identified because the adjacency matrix of type B nodes (blue nodes in Figure 2.1) is sparse, whereas systems in panel (a) and (d) cannot. The additional parameters' restrictions (condition (1b, 2a or 2b)) are due to an additional vector in the full rank condition (i.e. $E(G_{ab}y_b)$).

Proposition 1 has a natural interpretation in terms of instrumental variables. A multiple group framework adds an extra layer of exclusion restrictions. In fact, multiple sets of matrices provide additional instruments. The intuition is that when we distinguish nodes in different types, a higher number of possible

[16] For example, we can take advantage of linear independence of I_a, G_a and $G_{ab}G_{ba}$ (instead of I_a, G_a and G_a^2); and G_{ab} and G_aG_{ba}.

Figure 2.1: Identification with heterogeneous nodes



network intransitivities are formed. Appendix B provides technical details on the connection between identification in a single group model and a multiple group one.

2.4 The 2SLS estimator

Equation (2.2) cannot be consistently estimated by OLS because $G_a y_a$ and $G_{ab} y_b$ are correlated with ϵ_a.[17] We consider 2SLS estimation for the model in the spirit of Lee-Liu (2010). Following the standard technique used in spatial econometrics literature, we have the following optimal instruments from the two (symmetric) reduced forms

$$E(G_a y_a) = G_a(M_a(\phi_{ab} G_{ab} J_b E(B)\delta_b + E(A)\delta_a), \tag{2.6}$$

$$E(G_{ab} y_b) = G_{ab}(M_b(\phi_{ba} G_{ba} J_a E(A)\delta_b + E(B)\delta_a). \tag{2.7}$$

Recalling that $Z_a = [G_a y_a, G_{ab} y_b, E(A)]$ is a $n \times (k+2)$ matrix, we have $f_a = E(Z_a) = [E(G_a y_a), E(G_{ab} y_b), E(A)]$, so from equations (2.6) - (2.7) we have

[17] From equation (2.5), $G_a y_a = S_a(\phi_{ab} G_{ab} J_b B\delta_b + A\delta_a)$ where $S_a = G_a M_a$. OLS is not consistent because we have $E((G_a y_a)', \epsilon_a) = E((S_a(\phi_{ab} G_{ab} J_b B\delta_b + A\delta_a))', \epsilon_a) = E((S_a\epsilon_a)', \epsilon_a)$, since we assume that the $cov(\epsilon_a, \epsilon_b) = 0$ and $E(\epsilon_a) = E(\epsilon_b) = 0$. It follows that $E((S_a\epsilon_a)', \epsilon_a) = \sigma_a^2 Tr(S_a) \neq 0$. A similar argument holds for $G_{ab} y_b$.

$$Z_a = f_a + v_a = f_a + [(\phi_{ab} S_a G_{ab} J_b \epsilon_b + S_a \epsilon_a), (\phi_{ba} S_{ab} G_{ba} J_a \epsilon_a + S_{ab} \epsilon_b)][e_1, e_2]',$$

where e_1 is a first unit vector of dimension $(k+2)$, $S_a = G_a M_a$ and $S_{ab} = G_{ab} M_b$. These instruments are infeasible given the embedded unknown parameters. f_a can be considered a linear combination of IVs in

$$H_\infty^* = (S_a(G_{ab} J_b E(B), E(A)), S_{ab}(G_{ba} J_a E(A), E(B)), E(A)).$$

Furthermore, since $S_a = G_a M_a$ and $S_{ab} = G_{ab} M_b$ provided $\|\phi_a G_a\|_\infty + \|\phi_{ab}\phi_{ba} C_a\|\|_\infty \leq 1$ and $\|\phi_b G_b\|_\infty < 1$, we have $S_a = G_a \sum_{j=0}^{\infty} (\phi_a G_a + \phi_{ab}\phi_{ba} C_a)^j = G_a \sum_{j=0}^{\infty} (\phi_a G_a + \phi_{ab}\phi_{ba} G_{ab} \sum_{j=0}^{\infty} (\phi_b^j G_b^j) G_{ba})^j$. The same approximation holds for S_{ab}. It follows that

$$C_a = G_{ab} J_b G_{ba} = G_{ab}(\sum_{j=0}^{\infty} \phi_b^j G_b^j) G_{ba} = G_{ab}(\sum_{j=0}^{p} \phi_b^j G_b^j + (\phi_b G_b)^{p+1} J_b) G_{ba}.$$

This implies $\|C_a - \sum_{j=0}^{p} \phi_b^j G_b^j\|_\infty \leq \|(\phi_b G_b)^{p+1}\|_\infty \|C_a\|_\infty = o(1)\ as\ p \to \infty$.

$$S_a = G_a M_a = G_a \sum_{j=0}^{\infty} (\phi_a G_a + \phi_{ab}\phi_{ba} C_a)^j = G_a [\sum_{j=0}^{p} (\phi_a G_a + \phi_{ab}\phi_{ba} C_a)^j + (\phi_a G_a + \phi_{ab}\phi_{ba} C_a)^{p+1} S_a] \to \|S_a - \sum_{j=0}^{p} (\phi_a G_a + \phi_{ab}\phi_{ba} C_a)^j\|_\infty \leq \|(\phi_a G_a + \phi_{ab}\phi_{ba} o(1))^{p+1}\|_\infty \|S_a\|_\infty = o(1)\ as\ p \to \infty.$$

Hence, the approximation error by series expansion diminishes very quickly in a geometric rate, as long as the degree of approximation (p) increases as n increases. We can also replace S_a and S_{ab} by a linear combination. The instruments become

$$H_\infty^a = (G_a(I, G_a, G_a^2, \ldots (G_{ab}(I, G_b, G_b^2, \ldots) G_{ba})) \ldots (G_{ab}(I, G_b, G_b^2, \ldots) E(B), E(A)),$$

$$H_\infty^{ab} = (G_{ab}(I, G_b, G_b^2, \ldots (G_{ba}(I, G_a, G_a^2, \ldots) G_{ab})) \ldots (G_{ba}(I, G_a, G_a^2, \ldots) E(A), E(B)),$$

with an approximation error diminishing very quickly when K (or p) goes to infinity, where K denotes the number of instruments. Let us define $H_\infty = [H_\infty^a, H_\infty^{ab}, X_a^*, G_{ab}X_b]$ as the matrix of instruments and select an $n_a \times K$ sub-matrix H_K based on a p-order approximation of H_∞.[18] For instance, if we use the second order approximation of the infinite sums, $H_K = (H_2^a, H_2^{ab}, X_a^*, G_{ab}X_b)$ will be the first step best projector. The feasible 2SLS estimator for model (2.2) is

$$\hat{\mu} = (Z_a'\hat{P}_K Z_a)^{-1} Z_a'\hat{P}_K Y_a, \tag{2.8}$$

where $\hat{\mu} = (\phi_a, \phi_{ab}, \beta_a^*, \gamma_{ab})$ and $\hat{P}_K = H_K(H_K'H_K)^{-1}H_K'$.

2.4.1 Asymptotic Properties

This section derives the asymptotic properties of the many-instrument 2SLS estimator for heterogeneous network models. Cohen-Cole et al. (2012) and Liu (2013b) consider a network model with two endogenous variables and one adjacency matrix with multiple networks.[19] Our network model requires two endogenous variables, and two different adjacency matrices.[20] In Lee and Liu (2010) and Liu (2013b), the asymptotic approximation of the 2SLS estimator is based on many-instrument asymptotics, where the many instruments derive from the multiple network framework. In our model the many instruments derive from the (approximation of the) multiple adjacency-matrix framework. A multiple matrix framework results in an increasing number of instruments due to multiple approximations of the optimal instruments.[21] This complicates the derivations of the asymptotic properties of the many-instrument 2SLS estimator.

The following propositions establish the consistency and asymptotic normality

[18] Note that K is a function of the degree of approximations p.

[19] Kelejian and Prucha (2004) considers SAR models with multiple endogenous variables and a unique weights matrix.

[20] We consider the analysis with one network only. The extension to multiple networks extremely complicates the notation burden, but the theoretical results remain basically unchanged.

[21] See Section 2.4.

of the many-instrument 2SLS estimator in equation (2.8). Regularity conditions together with some discussion can be found in Appendix A. Some useful Lemmas are provided in Appendix B. All the proofs are listed in Appendix C. Let

$$F_a = \lim_{n\to\infty} \frac{1}{n} f_a' f_a,^{22}$$

$P_K S_a = \Psi_a$ and $P_K T_{ba} = \Xi_{ba}$, where $T_{ab} = S_{ab} G_{ba} J_b$.[23]

Proposition 2. Under assumptions 1-5, if $K/n \to 0$, then $\sqrt{n}(\hat{\mu} - \mu_0 - b) \xrightarrow{d} N(0, \sigma_a^2 F_a^{-1})$, where $b = (Z_a' P_K Z_a)^{-1}[e_1, e_2]\sigma_a^2[tr(\Psi_a), \phi_{ba} tr(\Xi_{ba})]' = O_p(K/n)$.

From Proposition 2, when the number of instruments K grows at a slower rate than the sample size n, the 2SLS estimator is consistent and asymptotically normal. However, the asymptotic distribution of the 2SLS estimator may not be centered around the true parameter value due to the presence of many-instrument bias of order $O_p(K/n)$ (see, e.g., Lee and Liu, 2010). We note that the leading order of the bias is the same as in Lee and Liu (2010) and Liu (2013b). However, the structure of the bias differs. Here, it depends on multiple approximations of the optimal instruments (see the beginning of Section 2.4). The condition that $K/n \to 0$ is crucial for the 2SLS estimator to be consistent. This appears evident if we look at the normal equation of our estimator: $\frac{1}{n} Z_a' P_K (Y_a - Z_a \hat{\mu})$. When $\hat{\mu} = \mu_0$ we have that $E(\frac{1}{n} Z_a' P_K (Y_a - Z_a \mu_0)) = [e_1, e_1]\sigma_a^2[tr(\Psi_a), \phi_{ba} tr(\Xi_{ba})]' = O_p(K/n)$ by Lemma B.2 in the Appendix. This converges to 0 only if the number of instruments grows more slowly than the sample size.[24] The following corollary characterizes different scenarios for different rates in which K diverges from n.

Corollary 2.1. *Under assumptions 1-5, (i) if* $K^2/n \to 0$, $\sqrt{n}(\hat{\mu} - \mu_0) \xrightarrow{d} N(0, \sigma_a^2 F_a^{-1})$;

[22] This is a crucial assumption. See the discussion in Appendix A after Assumption 4.

[23] To simplify the notation, we assume that $n \to \infty$ implies $n_a \to \infty$ and $n_b \to \infty$.

[24] Indeed, if we use a fixed number of instruments given by $\bar{H}$, the asymptotic distribution will be $\sqrt{n}(\hat{\mu} - \mu_0) \xrightarrow{d} N(0, \sigma_a^2 (\lim_{n\to\infty} \frac{1}{n} f_a' \bar{P} f_a)^{-1}$. Note that $(F_a - lim_{n\to\infty} \frac{1}{n} f_a' \bar{P} f_a) = \lim_{n\to\infty} f_a'(I - \bar{P}) f_a$, which is positive semi-definite in general. The 2SLS estimator with fixed number of instrument is generally not efficient. In order to have efficiency, we need to index our matrix of instruments with K and let K grow more slowly than the sample size.

(ii) if $K^2/n \to c < \infty$, $\sqrt{n}(\hat{\mu} - \mu_0) \xrightarrow{d} N(\bar{b}, \sigma_a^2 F_a^{-1})$, *where* $\bar{b} = \lim_{n\to\infty} \sqrt{n} b$.

The many-instrument bias of the 2SLS estimator can be corrected by the estimated leading-order bias (b) given in Proposition 2. Given consistent estimates of $\hat{\phi}_a$, $\hat{\phi}_b$, $\hat{\phi}_{ab}$, $\hat{\phi}_{ba}$, $\hat{\sigma}_a$ and $\hat{\sigma}_b$, the bias-corrected 2SLS estimator is

$$\hat{\mu_c} = (Z_a' P_K Z_a)^{-1} [Z_a' P_K Y_a - \hat{\sigma}_a^2 [e_1, e_2][tr(\Psi_a), \hat{\phi}_{ba}(\Xi_{ba})]']. \tag{2.9}$$

The following proposition shows that the bias-corrected estimator is properly centered around the normal distribution.

Proposition 3. Under assumptions 1-5, if $K/n \to 0$ and $\hat{\phi}_a$, $\hat{\phi}_b$, $\hat{\phi}_{ab}$, $\hat{\phi}_{ba}$, $\hat{\sigma}_a$ and $\hat{\sigma}_b$ are $\sqrt{n}-$consistent initial estimators, then $\sqrt{n}(\hat{\mu_c} - \mu_0) \xrightarrow{d} N(0, \sigma_a^2 F_a^{-1})$.

In the next subsection we discuss the case in which the number of endogenous variables (groups) grows with the sample size.

2.4.2 Estimation with Many Groups

So far, we have assumed that group numerosity does not depend on the sample size. We believe that, in practice, such an assumption is virtually always satisfied. For instance, if we increase the size of the sample, we will always have two genders: male and female. However, for completeness, it is interesting to explore whether having the number of groups growing together with the sample size affects the estimator properties.

In the many-instrument literature, Anatolyev (2013) and Imbens et al. (2011) have relaxed the assumption of a fixed number of *exogenous regressors*. To the best of our knowledge, the implications of relaxing the assumption of a fixed number of *endogenous regressors* have not been investigated yet.

Let us define g as the number of endogeneous variables and p as the degree of approximation (see Appendix C for an intuition of p as length of chains).

The following proposition characterizes the rate of divergence of g from n.

Proposition 4. if $K/n \to 0$, we have that $g = o(n^{1/p})$.

This means, that for our estimator to be consistent and asymptotically normal in this framework with many instruments and many endogenous variables we need g to grow more slowly than $n^{1/p}$.

For completeness, let us consider the link between the number of groups (i.e. endogenous variables) and the many-instrument asymptotics.

In our framework we have that $g/K \to 0$. In order to have a good performance of the estimator we need $K/n \to 0$. This implies $g/n = 1/s_g \to 0$, where s_g is the average size of groups. In words, in order to have a good performance of the estimators, we need the size of groups to be large enough. Furthermore, in order to have the estimator properly centered, we need $K^2/n \to 0$. This implies $g^2/n = g/s_g \to 0$. Therefore, for asymptotic efficiency, the average size of groups needs to be large enough compared to the number of groups. These results are similar to those in Lee and Liu (2010). However, the framework in Lee and Liu (2010) considers multiple networks embedded in a block-diagonal adjacency matrix (i.e. $G = diag(G_a, G_b)$) with the restriction that the within peer effects are the same for each network, (i.e. $\phi_a = \phi_b$) and there are no interactions between networks . If a network is defined as a group, then our framework can be considered as a generalization. We have different groups, with both within and between-group interactions. Our adjacency matrix is thus not block-diagonal.

2.4.3 Finite sample performance

In this section, we use simulation experiments to investigate the performance of the proposed estimator in small samples.

We conduct a Monte Carlo simulation study based on the following model

$$y_a = \phi_a G_a y_a + \phi_{ab} G_{ab} y_b + X_a^* \beta_a + G_{ab}^* X_b \gamma_{ab} + \epsilon_a,$$

$$y_b = \phi_b G_b y_b + \phi_{ba} G_{ba} y_a + X_b^* \beta_b + G_{ba}^* X_a \gamma_{ba} + \epsilon_b,$$

where X_a, X_b and $\epsilon \sim N(0,1)$. Borrowing from Lee and Liu (2010), we generate the G matrix as follows. First, for the i_{th} row of G, we generate an integer $d_i \in [0, 1, .., m]$ with a uniform probability function, where $m = 10, 20, 30$. Then we set the $(i+1)_{th}, \cdots, (i+d_i)_{th}$ elements of the i_{th} row of G to be ones. If $(i+d_i)_{th} < n_a$, the other elements in that row are zeros; otherwise, the entries of ones will be wrapped around such that the number of $d_i - n_a$ entries of the $i_t h$ row will be ones. We partition the matrix into four submatrices G_a, G_b, G_{ab} and G_{ba} with a random selection of rows and correspondent columns. The identifier variable used to select the two groups is generated by a Bernoulli distribution with p=0.5. The number of replications is 1000 and $n_a = n_b = 500$. We perform two experiments that are summarized in Table 2.1 and Table 2.2. Each column reports mean and standard error (in parenthesis) of the empirical distributions of different estimators. The first column shows 2SLS few IVs. It is based on equation (2.8) with the IV matrix H_K derived by the first order approximation of the best instruments (K=24). The second column reports the 2SLS many IVs, it is derived by the second order approximation of the best instruments (K=84). Finally, Column 3 shows the 2SLS bias-corrected. It is based on equation (2.9) with consistent estimates derived from the 2SLS few IVs.

Table 1 reports on the performance of the estimators when changing the density of the network, i.e. the number of connections. Each panel represents a different value of m, which indicates the maximum number of connections. The data are generated with $\beta_a = \beta_b = \gamma_a = \gamma_b = \gamma_{ab} = \gamma_{ba} = 0.5$. The peer effects parameters are set to: $\phi_a = \phi_b = 0.1$ and $\phi_{ab} = \phi_{ba} = 0.2$. The results show that all estimators perform well, with different nuances. In particular, one can observe the trade-off between bias and efficiency for the 2SLS many IVs when network density increases- the higher the density, the higher the gain in terms of efficiency with respect to the 2SLS few IVs. However, the bias (due to the

many instruments) increases as well. The bias correction that we propose is thus particularly beneficial when the network is dense.

Table 2.2 reports on the performance of the estimators when changing the heterogeneity within and between-group parameters. The simulation setup remains unchanged, but we now set the maximum number of connections to 20 and let the ϕ parameters vary. In the first panel, we consider $\phi_a = \phi_{ab} = \phi_b = \phi_{ba} = 0.1$ This is the benchmark framework in which peer effects are homogeneous. In the second panel, we introduce some heterogeneity in the within-group interaction effects. We set $\phi_a = \phi_b = 0.1$ and $\phi_{ab} = \phi_{ba} = 0.3$. In the third panel, peer effects are different both within and between groups. We set $\phi_a = 0.1$, $\phi_b = 0.2$ $\phi_{ab} = 0.4$ and $\phi_{ba} = 0.05$. Table 2 shows that the performance of the estimators does not depend on the values of the parameters- the ranking of the estimators in terms of efficiency and bias remains unchanged.

To test the robustness of our results, we have also performed two additional exercises.[25] First, instead of using randomly generated networks, we have used the Add Health's sociomatrix[26] as an adjacency matrix, thus replicating features of real-world social networks. Our aim is to understand whether the results of Table 2.1 are driven by the random generation of links. Second, we use uniform and gamma distributions to generate the errors of the data generating process. In doing so, our aim is to investigate whether and to what extent our i.i.d. assumption for the error terms in the derivation of large sample properties affects the finite sample Monte Carlo results. In both cases, the simulation results are very similar to those reported here.

[25] Results available upon request.

[26] A matrix derived from observed connections among students in the Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due to Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

2.5 Model Misspecification Bias

In this section, we investigate the bias occurring when the interaction structure is misspecified.

First, we analytically derive the bias that occurs when only within-group peer effects are considered, i.e. when interactions between groups are at work but ignored by the econometrician. We then use a simulation experiment to evaluate this bias in finite samples.

Second, we derive the mapping between the parameters of a model with homogeneous peer effects and those of a model with heterogeneous peer effects. We then use a simulation experiment to give an example of parameter mapping when peer effects are believed to be homogeneous but are actually heterogeneous in the data generating process (DGP).

Let us suppose the econometrician estimates the following model

$$y_a = (I - \phi_a G_a)^{-1}(X_a\beta_a + G_a^* X_a\gamma_a + \epsilon), \tag{2.10}$$

whereas the real DGP is

$$y_a = G_a(M_a(\phi_{ab}G_{ab}J_b B\delta_b + A\delta_a), \tag{2.11}$$

$$y_b = G_{ab}(M_b(\phi_{ba}G_{ba}J_a A\delta_b + B\delta_a). \tag{2.12}$$

This model misspecification results in an estimator of the endogenous effect ϕ_a that is inconsistent. First, we are omitting the influence of the outcome of type B agents. Second, we do not consider the indirect connections among type A nodes passing through type B nodes. As a result, G_a^k, with $k \geq 2$, is misspecified. Therefore, the commonly used instrument G_a^2 might not be valid as the exclusion restrictions might be violated. Third, we misspecify the contextual effects ($G_a^* X_a$)

by ignoring the characteristics of other-type peers. [27]

Analytically, the bias is

$$E(\hat{\mu_a}) = \mu_a + E(Z_a' P_a Z_a)^{-1} Z_a P_a (\phi_{ab} G_{ab} y_b + G_{ab}^* X_b \gamma_{ab}),$$

where $Z_a = [G_a y_a, X_a, G_a^* X_a]$ and $\mu_a = (\phi_a, \beta_a, \gamma_a)$. The bias is positively correlated with the direct influence of type B on type A, as captured by the peer effects from B to A and the influence of the characteristics of type B on type A.

Table 2.3 shows the extent of this bias in finite samples through a Monte Carlo simulation. Table 2.3 represents the performance of the 2SLS few IVs following the same experiment design as in the previous section.[28] We report on the case in which the maximum number of connections is 10 for each node (as in panel 2 of table 2.1).[29] The first column reports the real value of the parameters. The second column shows the performance of the 2SLS estimator in the misspecified model. When interactions between groups are at work but ignored by the econometrician it results in the size of the bias derived above. The third column shows the results of the estimator when the econometrician considers the correct DGP (equations (2.11) and (2.12)), but does not use the approximation of optimal instruments (in equation (2.8)). In other words, we consider the case where the traditional network IV approach is applied mechanically, thus $G_a^2 X_a$ and $G_{ab} G_b X_b$ are used as instruments respectively for $G_a Y_a$ and $G_{ab} Y_b$. In short, only within-group instruments are considered. The resulting 2SLS estimator is consistent but not efficient. The fourth column reports the performance of our 2SLS few IVs (in equation (2.8)), which considers the H_k matrix derived in Section 2.4(i.e which

[27] This issue also arises when full information about node characteristics and network structure is not available. See Chandrasekhar and Lewis (2011a), Liu et al. (2013) and Liu (2013a) for problems related to the use of sampled network data.

[28] We use the 2SLS few IVs to ease the comparison of 2SLS estimators with the misspecified set of instruments. Observe that the bias considered here is due to the misspecification of the model rather than to the many-instrument issue.

[29] The simulation results in the other cases, i.e. when the maximun number of simulations is 20 or 30, are very similar.

also includes between group instruments).[30] Mean values for each coefficient's empirical distribution and standard errors (in parenthesis) are reported.

Table 2.3 shows that the bias is large in the second column, especially for the β coefficients. In the second column the bias is not large, but the problem is efficiency. Our approach (third column) reveals no bias and improved efficiency.

In our second exercise, we consider the case in which the econometrician estimates a standard network model (model (2.1)) when the real DGP is characterized by heterogeneous peer effects (model (2.11) - (2.12)).

Let us define the following $n \times n$ matrices

$$G(a) = \begin{bmatrix} G_a & O_{ab} \\ O_{ba} & O_b \end{bmatrix}, \quad G(ab) = \begin{bmatrix} O_a & G_{ab} \\ O_{ba} & O_b \end{bmatrix},$$

$$G(ba) = \begin{bmatrix} O_a & O_{ab} \\ G_{ba} & O_b \end{bmatrix}, \quad G(b) = \begin{bmatrix} O_a & O_{ab} \\ O_{ba} & G_b \end{bmatrix},$$

where O_l is a $l \times l$ matrix of zeros and O_{lk} is a $l \times k$ matrix of zeros. Let us suppose for simplicity that $\beta = \beta_a = \beta_b$ and $\gamma = \gamma_a = \gamma_{ab} = \gamma_{ba} = \gamma_b$ and focus our attention on the peer effects parameters. In this case model (2.1) can be written as

$$\begin{aligned} Y &= \phi GY + X\beta + G^*X\gamma + \epsilon \qquad (2.13) \\ &= (\phi_a G(a) + \phi_{ab} G(ab) + \phi_{ba} G(ba) + \phi_b G(b))Y + X\beta + G^*X\gamma + \epsilon. \end{aligned}$$

Hence, the peer effects parameter, ϕ, is the following non-linear function of heterogeneous peer effects

$$\phi = \phi_a G^{-1}G(a) + \phi_{ab} G^{-1}G(ab) + \phi_{ba} G^{-1}G(ba) + \phi_b G^{-1}G(b).$$

[30] First order approximation of optimal instruments is considered.

If $\phi_a = \phi_b = \phi_{ab} = \phi_{ba} = \phi$, then

$$\phi_a G(a) + \phi_{ab} G(ab) + \phi_{ba} G(ba) + \phi_b G(b) = \phi(G(a) + G(ab) + G(ba) + G(b)) = \phi G.$$

Table 2.4 contains the results of a simulation experiment in which we estimate model (2.13), for different values of ϕ_a, ϕ_b, ϕ_{ab} and ϕ_{ba}. The simulation set-up is as before- the data generating process remains as in equations (2.11) and (2.12)). The estimator considered is the 2SLS few IVs.

In the first column, we set all the ϕ parameters equal to 0.1. In fact, the 2SLS few IVs consistently estimates $\phi = \phi_a = \phi_b = \phi_{ab} = \phi_{ba}$. In the second column, we add some heterogeneity. We set $\phi_{ab} = 0.3$ and $\phi_{ba} = 0.3$, leaving the other parameters unchanged. The third column corresponds to the case in which all the ϕ parameters are different. As expected, as soon as some heterogeneity is introduced, the estimated value of ϕ is not informative at all.

2.6 Impact Evaluation and Treatment Effect

Let us now highlight the importance of our analysis for the identification of treatment response with spillovers. Let A be the set of *eligible* recipients and B the set of *ineligible* recipients of a treatment (respectively *eligibles* and *ineligibles* hereafter). The treatment is administrated using a *randomized controlled experiment.* Having in mind policy interventions such as conditional cash transfer or microfinance subsidies can be useful. Let T_a be the binary treatment vector whose i_{th} element is $T_{a,i} = \{0, 1\}$, which indicates whether i is treated or not (among the

eligibles).[31] Model (2.2) and (2.3) can be written as

$$Y_a = \phi_a G_a Y_a + \phi_{ab} G_{ab} Y_b + X_a^* \beta_a^* + \delta_a T_a + \rho_a G_a T_a + G_{ab}^* X_b \gamma_{ab} + \epsilon_a, \tag{2.14}$$

$$Y_b = \phi_b G_b Y_b + \phi_{ba} G_{ba} Y_a + X_b^* \beta_b^* + G_{ba}^* X_a \gamma_{ba} + G_{ba} T_a \rho_{ba} + \epsilon_b. \tag{2.15}$$

In this model, the Stable Unit Treatment Value Assumption (SUTVA)[32] doesn't hold because (i) spillovers are at work and (ii) spillovers are heterogeneous. To the best of our knowledge, there are no studies that consider violations of the SUTVA because of (ii). This is surprising given that heterogeneity in spillovers is naturally implied by differences between eligibles and ineligibles.

Our results in Sections 2.4 provide consistent and efficient estimators for the parameters of model (2.2) - (2.3).[33]

2.6.1 Average Treatment Effect with Heterogeneous Spillovers

The Average Treatment Effect in our context can be written as[34]

[31] Our analysis can be easily adapted to the case of continuous or multinomial treatment. It is also useful to recall an assumption already listed in the previous sections for estimator properties, $G \perp T_a$, which here states that the treatment does not change the network topology. This assumption relates to Manski (2013) which assumes that reference groups are person-specific and treatment-invariant (unable to be manipulated by the policy maker).

[32] Following Rubin (1986), SUTVA states that potential outcomes depend on the treatment received, and not on what treatments other units receive and that there are no "hidden treatments".

[33] As mentioned in the Introduction, we do not consider *direct* treatment effect heterogeneity. This assumption can be relaxed, allowing for a double form of heterogeneity: one coming from individual characteristics, the other from the interactions. The identification becomes much more complex. We leave this extension for future research. Following Manski (2013), we also assume here that the treatment does not change the network topology, i.e. that the policy maker cannot manipulate reference groups.

[34] When the treatment is a randomized control experiment, the average treatment effect is equal to the average treatment effect on treated.

$$ATE = E(Y_i|i \in A,\ T_{a,i} = 1, X, G) - E(Y_i|i \in A,\ T_{a,i} = 0, X, G). \tag{2.16}$$

From the reduced form of equation (2.14)

$$ATE = \delta_a E_{E_T}(m_{a,ii}), \tag{2.17}$$

where $m_{a,ii}$ is the ii_{th} element of M_a and $E_{E_T}(\cdot) = E(\cdot|i \in A, T_{a,i} = 1, X, G)$ indicates the expected value over the treated eligibles. The Average Treatment Effect is thus equal to the direct impact of the treatment on the individual i (i.e. δ_a) plus the indirect effect of other agents' spillovers on i triggered by i's treatment (but not triggered by other nodes' treatment)

$$\delta_a M_a = \delta_a I_a + \delta_a \sum_{k=1}^{\infty} (\phi_a G_a + \phi_{ab}\phi_{ba} C_a)^k.$$

Observe that $m_{a,ii}$ is a function of $(G_a, G_b, G_{ab}, G_{ba}, \phi_a, \phi_b, \phi_{ab}, \phi_{ba})$. This implies that when network interactions are at work, the ATE depends on network topology and strength of outcome spillovers among agents. As a result, an individual can have a high increase in outcome even if she has a low treatment direct impact (a low δ_a) but she is *central* in the network.[35] Observe that even if $\delta_{a,i} = \delta_a$(i.e. the treatment effect is homogenous) the ATE can be heterogeneous because of the different position of nodes in the network. Indeed, the ATE can be decomposed into two parts

$$ATE = \underbrace{\delta_a}_{\text{DTE}} + \underbrace{\delta_a E_{E_T}[(diag(M_a - I_a)]}_{\text{FLTE}}. \tag{2.18}$$

The first part is the *Direct Treatment Effect* (hereafter DTE), whereas the second part is the effect of the treatment due to the interactions among agents, i.e.

[35] Of course the centrality itself is not a sufficient condition, a high level of spillovers is required.

the effect of i's treatment that impact i through other nodes. We denote the latter effect as *Feedback Loop Treatment Effect* (hereafter $FLTE$). The sample counterpart of equation (2.18) is

$$\hat{ATE} = \mu_t'[\hat{\delta}_a diag(\hat{M}_a)]\mu_t \frac{1}{n_a^t} = \hat{\delta}_a \frac{1}{n_a^t} \sum_{i \in N_a^T} \hat{m_{a,ii}}, \tag{2.19}$$

where N_a^T is the set of treated individuals which has cardinality $n_a^t < n_a$, μ_t is the $n_a^t \times 1$ selector vector for that units and $\hat{M}_a = M_a(\hat{\phi_a}, \hat{\phi_b}, \hat{\phi_{ab}}, \hat{\phi_{ba}})$ is the estimated counterpart of M_a.

Treatment Effect Misinterpretation and Bias When SUTVA holds, $ATE = DTE$. If interferences are at work, then $ATE \neq DTE$. However, the problem is not only about interpretation. We show below that if spillovers are ignored, then the parameter estimates can be inconsistent. Suppose that a treatment is administered to $n_a^t < n_a$ subjects and we ignore interactions among them. Estimation of δ_a is based on the following regression

$$Y_a = X_a\beta_a + T_a\delta_a + \epsilon_a^*, \tag{2.20}$$

where $\epsilon_a^* = \rho_a G_a T_a + \phi_a G_a Y_a + \phi_{ab} G_{ab} Y_b + \epsilon_a$ contains the three relevant spillover effects omitted:[36] *(i)* the direct treatment spillover from other eligibles $\rho_a G_a T_a$, *(ii)*the endogenous outcome spillover from other treated eligibles $\phi_a G_a Y_a$ and *(iii)* the endogenous outcome spillover from ineligibles $\phi_{ab} G_{ab} Y_b$. Misinterpretation occurs because the estimate of δ_a is interpreted as a DTE while, if the data generating process is given by equations (2.14) and (2.15), it is an ATE. Bias can occur if the treatment is correlated with the three components listed above

[36] The other omitted terms, $X_a^*\beta_a^*$ and $G_{ab}^* X_b \gamma_{ab}$, are independent from the treatment.

$$\hat{\delta_a} = \delta_a + bias = \delta_a + (T_a'T_a)^{-1}T_a \quad \{ \quad \rho_a G_a T_a$$
$$+ \quad \phi_a G_a M_a[\phi_{ab} G_{ab} J_b(\rho_{ba} G_{ba} T_a) + T_a \delta_a + \rho_a G_a T_a]$$
$$+ \quad \phi_{ab} G_{ab} M_b[\phi_{ba} G_{ba} J_a(\delta_a T_a + \rho_a G_a T_a) + \rho_{ba} G_{ba} T_a]\}.$$

The bias is due to the spillover effects coming from the three omitted components listed before. By correctly specifying the interaction structure we can consistently estimate the direct treatment effect purged of the influence of the three omitted components.

It should appear clear from our discussion that, if the spillovers' coefficients and the direct treatment effect are positive, neglecting between and within-group interactions result in an overestimation of the direct treatment effect. Manski (2013) defines this scenario as *Reinforcing Interactions*. Of course one can imagine different scenarios where interactions are not reinforcing and, on the contrary, are *Opposing Interactions.*

Our approach has an advantage from this point of view- it allows interactions between and within groups to be heterogeneous (e.g. *Reinforcing Interactions* within groups members and *Opposing Interactions* between groups). We also note that, using again Manski (2013)'s terminology, our framework can be adapted to the estimation of social interaction with *leaders* and *followers*, labeling those agents as groups A and B.

2.6.2 Indirect Treatment Effect

As mentioned before, the Indirect Treatment Effect (hereafter ITE) has been an object of interest in several papers. Most of the existing papers focus attention on the indirect effect on ineligibles (see, e.g. Angelucci and De Giorgi, 2009). However, when the population is split into two sets, it is also natural and

interesting from a policy perspective to understand whether different types of untreated (eligible or ineligible), are differently impacted by the treatment. Let us define $ITEE$ and $ITEI$ as the Indirect Treatment Effect on Eligibles and the Indirect Treatment Effect on Ineligibles, respectively.

The Indirect Treatment Effect on Eligibles in our model is

$$ITEE = E(Y_i|i \in A, M_iT \neq 0 \cap T_{a,i} = 0, X, G) - E(Y_i|i \in A, M_iT = 0 \cap T_{a,i} = 0, X, G), \tag{2.21}$$

whereas the Indirect Treatment Effect on Ineligibles can be defined as

$$ITEI = E(Y_i|i \in B, M_iT \neq 0, X, G) - E(Y_i|i \in B, M_iT = 0, X, G), \tag{2.22}$$

where M_i is the i_{th} row of $M = \left(\begin{bmatrix} I_a \\ I_b \end{bmatrix} - \begin{bmatrix} G_a\phi_a & G_{ab}\phi_{ab} \\ G_{ba}\phi_{ba} & G_b\phi_b \end{bmatrix} \right)^{-1}$, $T = [T_a, 0_b]$, and 0_b is a $n_b \times 1$ vector of zeros. $M_iT = 0$ indicates that i is not affected by any of the treated nodes (i.e. that there are no direct and indirect paths in the networks between i and a treated node).

Let us now suppose that, given our data generating process (equations (2.14) and (2.15)) we are asked by a policy maker to evaluate the Indirect Treatment Effects after a treatment administered to the eligibles (i.e. to a subset of A). From model (2.14) - (2.15) we can derive the following formulas

$$ITEE = E_{E_u}[M_{a_i}(\phi_{ab}G_{ab}J_b(\rho_{ba}G_{ba}T_a) + \delta_aT_a + \rho_aG_aT_a)],$$

$$ITEI = E_I[M_{b_i}(\phi_{ba}G_{ba}J_a(\delta_aT_a + \rho_aG_aT_a) + \rho_{ba}G_{ba}T_a)],$$

where M_{a_i} is the ith row of M_a, M_{b_i} is the ith row of M_b , $E_I(\cdot) = E_I(\cdot|i \in$

$B, M_iT \neq 0, X, G)$ indicates the expected value over the (indirectly treated) ineligibles, and $E_{E_u}(\cdot) = E(\cdot|i \in A, M_iT \neq 0 \cap T_{a,i} = 0, X, G)$ indicates the expected value over (indirectly treated) untreated eligibles. Observe that these estimands depend on direct and indirect connections because of network-based spillovers. More formally, they can be decomposed into different parts. For instance, $ITEE$ may be decomposed into three effects. The first term , $\delta_a M_a$, captures propagation of the treatment via outcome spillovers.[37].

The nice feature of this derivation of $ITEI$ and $ITEE$ is that instead of simply addressing the question whether an ITE is different from zero, we can also decompose it into different sources of treatment's transmission. For instance, one can find that the treated population has a strong reaction to the treatment (δ_a and ρ_a are high) and transmits it to ineligibles through low magnitude peer effects (ϕ_{ab} is low). The same level of $ITEI$, however, can also arise from a scenario where there is a low reaction to the treatment within group (δ_a and ρ_a are low) and a large transmission between groups (ϕ_{ab} is high).

Understanding these different channels is paramount for policy purposes. Most importantly, our framework enables the researcher to distinguish the role of *contextual effects* from *peer effects* in transmitting the treatment. In other words, one can quantify how much of the effect is generated by the direct effect of the treatment through exogenous variables (as captured by δ_a, ρ_a and ρ_{ba}) and how much is due to spillovers through outcomes (as captured by ϕ_{ba}, ϕ_a, ϕ_b and ϕ_{ab}). Note also that having these estimates at hand, one can understand which effects (within eligibles, within ineligibles and between them) are the dominant ones in

[37]Given that $M_a = (I - \phi_a G - \phi_{ab}\phi_{ba}C_a)^{-1}$, we have $M_a\delta_a = I_a\delta_a + [(I_a - \phi_a G - \phi_{ab}\phi_{ba}C_a)^{-1} - I]\delta_a$. The first term is the diagonal matrix of treatment direct effects which has (by definition) no impact on the untreated, while the second term represents the propagation of those effects through the network via endogenous spillovers (i.e. changes in outcomes due to treatment). The second term, $\rho_a M_a G_a$, measures the spillover arising from the treatment given to other units ($\rho_a G_a$), as well as its amplification through interactions (as captured by M_a). Finally, $\phi_{ab}\rho_{ba}M_a G_{ab} J_b G_{ba}$, denotes the spillover accruing to ineligibles distinguished between outcome amplification ($M_a G_{ab} J_b$) and (indirect) treatment effect ($\rho_{ba} G_{ba}$). A similar decomposition can be applied to $ITEI$

spreading out the policy's beneficial effect.

We can thus simply use the sample counterpart to estimate the $ITEE$ and $ITEI$

$$\hat{ITEE} = \mu_u'[\hat{M}_a(\hat{\phi}_{ab}G_{ab}\hat{J}_b(G_{ba}\hat{\gamma}_{ba}) + \hat{\beta}_a + G_a\hat{\gamma}_a)]\mu_u\frac{1}{n_a^u},$$

$$\hat{ITEI} = \iota_b'[\hat{M}_b(\hat{\phi_{ba}}G_{ba}\hat{J}_a(\hat{\beta}_a + G_a\hat{\gamma_a}) + G_{ba}\hat{\gamma_{ba}})]\iota_a\frac{1}{n_b},$$

where $n_a^u < n_a$ is the number of eligibles who are untreated, μ_u is the $n_a \times 1$ selector vector for that units and ι_l is an $n_l \times 1$ vector of ones.

2.6.3 Total Treatment Effect

One can also be interested in evaluating the treatment effect on the entire population (or network). As the SUTVA has been removed and spillovers are in place, it is useful to derive the Total Treatment effect (hereafter TTE). Following our previous notation we have the following definition for TTE

$$TTE = E(Y_i|i \in A \cup B, M_iT \neq 0, X, G) - E(Y_i|i \in A \cup B, M_iT = 0, X, G).$$

This represents the treatment effect on a generic individual in the network (eligible or ineligible). Its sample counterpart is

$$\hat{TTE} = \iota' \left(\begin{bmatrix} I_a \\ I_b \end{bmatrix} - \begin{bmatrix} G_a\hat{\phi}_a & G_{ab}\hat{\phi_{ab}} \\ G_{ba}\hat{\phi_{ba}} & G_b\hat{\phi}_b \end{bmatrix} \right)^{-1} \left(\hat{\delta}_a \begin{bmatrix} T_a \\ O_b \end{bmatrix} + \begin{bmatrix} G_a\hat{\rho}_a & G_{ab} \\ G_{ba}\hat{\rho}_{ba} & G_b \end{bmatrix} \begin{bmatrix} T_a \\ O_b \end{bmatrix} \right) \iota\frac{1}{n},$$

where ι is an $n \times 1$ vector of ones. Note that the $\hat{TTE}$ is basically the weighted average of $\hat{ATE}$, $\hat{ITEE}$, and $\hat{ITEI}$.

2.6.4 Control Group

It is well-known that the ATE, $ITEE$, $ITEI$ and TTE are identified if we have a control group, i.e. if we can distinguish sample of units who are not treated (directly or indirectly). This can be quite challenging when estimating the indirect treatment effects. In a network context, we have two possibilities: (i) a *multiple network-based approach* and (ii) *topology-driven approach.*

In the first case, we have multiple networks, some of which are randomly treated and others which are not- offering a valid control group. A similar scheme is often followed for policy design and evaluation in a non-network context.[38]

The second possibility is unique to a network approach and exploits the architecture of networks. When information on actual connections is available and the *direction* is known, it may be possible to estimate ATE, $ITEE$, $ITEI$ and TTE using only one network. The network topology determines the possibility of having the control population if there are some nodes in the network that are not influenced by a treatment to other nodes. For example, let us consider the network in Figure 2, where the red nodes are treated and the blue ones are not. According to the directions of the edges (arrows in the picture), the blue node i is influenced by red nodes whereas the blue node j is not. Therefore, the direction of the links between nodes stemming from this network topology allows us to distinguish between indirectly treated nodes (node i) and control group nodes (node j).[39]

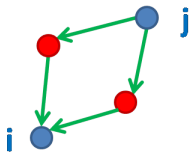
2.6.5 Policy Experiments

Manski (2013) studies treatment response in settings with endogenous effects. In this framework, endogenous effects are seen as a mechanism in which the treat-

[38]For example, in PROGRESA, a set of treated and untreated villages are surveyed (see Angelucci and De Giorgi (2009) for more details on the program design).

[39]Note also that we need these two kinds of nodes to be comparable in terms of characteristics.

Figure 2.2: *Topology-driven* policy evaluation design



ments could propagate.[40] The main objects of interest are $P[Y]$ and $P[Y(T_a)]$, the outcome distributions respectively without and with a treatment T_a administered to the population. Policy makers are usually interested in comparing these two distributions since interventions are often finalized to reduce inequality between a disadvantaged cluster and the rest of the population. The marginal effect of T on Y accounts for the adjustment of the outcome after a policy intervention.

In this paper, we consider a network framework with heterogeneous peer effects similar to Manski (2013).[41] In this section, we numerically study the empirical density functions $P[Y_a] - P[Y_a(T_a)]$ and $P[Y_b] - P[Y_b(T_a)]$, where the subscripts indicate the reference to eligible and ineligible populations.

We perform a numerical simulation to asses the extent to which the underlining heterogeneity of the endogenous effects can affect the outcome response for different groups. Our goal is to provide evidence about the individual and aggregate implications of this heterogeneity. In the simulation experiment below, we show that for some values of ϕ_a, ϕ_b, ϕ_{ba} and ϕ_{ab} it may be (paradoxically) more convenient to treat a group other than the target one. This has implications for the study of socio-economic inequality. Importantly, by allowing estimation of all the different parameters of interest, our model specification can be used to understand what nodes (or which type of nodes) should be targeted by a social planner whose final goal is to maximize an aggregate outcome or to converge to

[40] If a dynamic model is at work, then a social multiplier may also arise in terms of expectations.

[41] The framework in Manski (2013) considers only one group, thus homogeneous peer effects (and no between group interactions).

a desired distribution of individual outcomes.

We present an experiment where we treat a random sample of nodes and simulate the treatment's propagation through a network characterized by *heterogeneous peer effects*. More specifically, we look at the increase of type A and type B nodes'outcomes once a certain set of nodes receives a treatment[42].

Two exercises are implemented. In the first, we evaluate aggregate effects, i.e. the change in the sum of outcomes (for both type A and type B individuals) which follows a treatment for different values of peer effects (i.e. ϕ_a, ϕ_b, ϕ_{ba}, ϕ_{ab}). In the second exercise, we look at distributional effects, i.e. at changes in the empirical distribution of individual node outcomes for different sets of peer effects parameters following the policy intervention .

Figures 2.5 and 2.6 report on the first exercise. Figure 2.5 depicts the results when fixing $\phi_b = \phi_{ba} = 0.1$ and varying ϕ_a and ϕ_{ab}. We generate a grid of values for parameters resulting from two sequences: $\phi_a = 0.02, 0.04, ..., 0.50$ and $\phi_{ab} = 0.02, 0.04, ..., 0.50$. For each couple (ϕ_a, ϕ_{ab}) we generate one hundred independent replications using the same DGP as described in Section 2.4.3 and compute $Y_a^s = \sum_{i \in A} y_i$ and $Y_b^s = \sum_{i \in B} y_i$. We then select a random sample of one hundred type A nodes to be treated. This treatment is represented by an $n_a \times 1$ vector T_a of zeros for non treated nodes and ones for treated nodes. Finally, we compute $Y_a^{s*} = \sum_{i \in A} y_i^*$ and $Y_b^{s*} = \sum_{i \in B} y_i^*$, where $y_i^* = y_i + \frac{\partial y_i}{\partial T_a} T_A$. This exercise represents the case where group A nodes are treated and there are low interactions

[42] We compute the marginal effect matrix of T_a on Y_a multiplied by the treatment vector

$$\frac{\partial E(Y_a|G,X)}{\partial T_a} T_a = M_a(\phi_{ab} G_{ab} J_b(\rho_{ba} G_{ba} T_a) + \delta_a T_a + \rho_a G_a T_a).$$

Note that when there are no interactions between the two groups (or only type A nodes are considered in the analysis), we have $\frac{\partial E(Y_a|G,X)}{\partial T_a} T_a = S_a(\delta_a T_a + \rho_a G_a T_a)$, where $S_a = (I - \phi_a)^{-1}$. This is the marginal effect matrix in a standard peer effects model.

The marginal effect matrix of T_a on Y_b is

$$\frac{\partial E(Y_b|G,X)}{\partial T_a} T_a = M_b(\phi_{ba} G_{ba} J_a(\delta_a T_a + \rho_a G_a T_a) + \rho_{ba} G_{ba} T_a). \tag{2.23}$$

Observe that the marginal effect of T_a on Y_a is different from the marginal effect of T_a on Y_b -an increase in T_a differently affects nodes depending on their type.

between nodes A and nodes B ($\phi_{ba} = 0.1$).

From equations (2.14) and (2.15) we have

$$\Delta y_i = y_i^* - y_i = \frac{\partial y_i}{\partial T_a} T_a = \begin{cases} M_{a_i}(\phi_{ab} G_{ab} J_b(\rho_{ba} G_{ba} T_a) + (\delta_a + G_a \rho_a) T_a) & if\ i \in A \\ M_{b_i}(\phi_{ba} G_{ba} J_a(\delta_a + \rho_a G_a) T_a + \rho_{ba} G_{ba} T_a) & if\ i \in B \end{cases}.$$

Figure 2.5 represents the differences $\Delta Y_a^s = Y_a^{s*} - Y_a^s = \sum \frac{\partial Y_a}{\partial T_a} T_a$ and $\Delta Y_b^s = Y_b^{s*} - Y_b^s = \sum \frac{\partial Y_b}{\partial T_a} T_a$ for all the possible combinations (ϕ_a, ϕ_{ab}).[43]

Figure 2.5 shows that ΔY_a^s increases steadily with ϕ_a (and slightly with ϕ_{ab}), whereas ΔY_b^s remains roughly unchanged. These results are not surprising. If there are no interactions (or low interactions) between the two groups, then there is no reason why the outcome of group B should change. The variation in the outcome of the group A depends on the extent of the endogenous effects (ϕ_a). If instead there are interactions between the two groups, then the treatment response depends on both ϕ_a and ϕ_{ab}. For example, assuming a positive effect of the independent variable, if a policy intervention targets a group when the two groups have the same outcome profile, we expect an increase in *inequality* in terms of outcomes between the two types when the within-peer effects (ϕ_a) are high and the between-peer effects (ϕ_{ab}) are low.

Figure 2.6 depicts the results when fixing $\phi_b = \phi_{ab} = 0.1$, and varying ϕ_a and ϕ_{ba}. The experiment design remains unchanged. This exercise represents the case where group A nodes are treated and there are increasing influences within nodes A and from nodes A and nodes B (ϕ_{ba} increasing up to 0.5). Figure 2.6 shows that an increase of ϕ_{ba} is beneficial for ΔY_b^s, as type B nodes receive an impulse from type A nodes. Interestingly, type B nodes may actually benefit even more than A nodes (the treated group). Our results shows that when $\phi_{ba} > 0.20$, we observe $\Delta Y_b^s > \Delta Y_a^s$. In terms of policy effects, this means that if a policy targets one

[43] Some combinations are missing in the grid because it is unlikely to draw G_a and G_{ab} such that $\|\phi_a G_a\|_\infty + \|\phi_{ab}\phi_{ba} C_a\|\|_\infty \leq 1$. These combinations are at the edge of the parameter space.

group but peer effects between groups are high, then we can observe increasing *inequality* between the two groups, rather than the expected decrease (assuming that the targeted group has a lower starting outcome). In terms of the estimands derived in Section 2.6 , note that the blue surfaces in Figures 2.5 and 2.6 are simply $ITEI \times n_b$ while the red ones are $ITEE \times n_a^u + ATE \times n_a^t$, plotted for different combinations of parameters.

In the second exercise, we consider four points from the grid formed by ϕ_a and ϕ_{ba} and look at the empirical distributions of $\Delta y_{i\in A}$ and $\Delta y_{i\in B}$. We estimate these distributions using a normal kernel density. We consider the case where $\phi_a = 0.1$ and $\phi_{ba} = 0.1$ as a benchmark and then increase the strength of peer effects among agents in different ways.

In Figure 2.7 we increase the effect within group A only ($\phi_a = 0.3$.). While this change is irrelevant for type B nodes, it has interesting implications for the distribution of outcomes among type A nodes (Panel a). While in the benchmark model (the single line), the distribution is quasi-bimodal (due to the treated and non-treated A nodes), an increase of ϕ_a smooths the distribution (the bold line). In other words, the higher the endogenous effects, the more evenly the benefits of the policy intervention are shared among nodes (individuals).

In Figure 2.8 we increase the between-group effect only ($\phi_{ba} = 0.3$). Type A density remains basically unchanged (Panel a) . The impact is instead apparent on the outcome distribution of type B nodes (Panel b). One can observe an important shift to the right. This means that non treated type B nodes benefit more than non-treated type A nodes (from the treatment to type A nodes).

The red and blue curves in Figures 2.7 and 2.8 are the empirical density functions $P[Y_a(T_a) - Y_a]$ and $P[Y_b(T_a) - Y_b]$, respectively. They have $\frac{ITEE \times n_a^u + ATE \times n_a^t}{n_a}$ and $ITEI$ as expected values.

2.7 Concluding Remarks

We generalize the linear-in-means model to the presence of two groups and between-group interactions. We derive the sufficient conditions to identify the model and propose efficient 2SLS estimators. We characterize the bias which arises when interactions are ignored and evaluate it in finite sample using simulation experiments. We illustrate the relevance of these issues for policy purposes. If peer effects are seen as a mechanism in which the treatments could propagate through the networks, then accounting for *heterogeneous peer effects* and *between-group interactions* can be helpful in designing and evaluating policy interventions that alter the outcome distribution. We show that when between-group interactions are strong, an impulse to a given group can engender benefits to another group which are even higher than those accruing to the target group. Examples of types of interventions where the local non-target population may also be indirectly affected by the treatment through social and economic interaction with the target population are widely varied. For example, the recipients of conditional cash transfers may share resources with ineligible households who live in the same community or with extended family members, which could affect the incentives to accumulate human capital (Angelucci et al. (2010)). School vouchers or other incentives (such as equipment provision) to increase schooling of indigent children may increase the learning ability of untreated children if, for example, textbooks or computers are shared among classmates. A number of organizations promote the deworming of children in the developing world as a public health and development strategy. Supplying deworming drugs to a group of children may benefit untreated children by reducing disease transmission, thus lowering infection rates for both groups.

In sum, our paper contributes to the literature by providing a framework able to decompose the treatment response into different components, including the crucial difference between endogenous effects and effects stemming from exoge-

nous variations in the characteristics of the treated.

APPENDIX

Appendix A: Assumptions and Discussions

Let us introduce some notation and assume the following regularity conditions: a sequence of square matrices $\{A\}$, where $A = [A_{ij}]$, is defined "uniformly bounded in absolute value" (UB) if there exists a constant $c_b < \infty$ (that does not depend on n) such that $\|A\|_\infty = \max_{i=1,\cdots,n} \sum_{j=1}^{n} |A_{ij}| < c_b$ and $\|A\|_1 = \max_{j=1,\cdots,n} \sum_{i=1}^{n} |A_{ij}| < c_b$. We indicate that $\{A\}$ is bounded only in row (column) sum absolute value as UBR (UBC). For the sake of simplicity we will assume that $n \to \infty$ implies $n_a \to \infty$ and $n_b \to \infty$.

Assumption 2.2. *The elements of ϵ_a and ϵ_b are iid with zero mean, variance σ_a^2 and σ_b^2 respectively, and zero covariance. Moments higher than the fourth exist.*

Assumption 2.3. *The elements of X_a and X_b are uniformly bounded constants, X_a and X_b have full rank k, and $\lim_{n_a \to \infty} \frac{1}{n_a} X_a' X_a$ and $\lim_{n_b \to \infty} \frac{1}{n_b} X_b' X_b$ are finite and non singular.*

Assumption 2.4. *The sequences of matrices $\{G_a\}$, $\{G_{ab}\}$, $\{G_b\}$, $\{G_{ba}\}$, $\{M_a\}$, $\{M_b\}$, $\{J_b\}$, and $\{J_a\}$ are UB.*

The first assumption is needed in order to apply the Kelejian and Prucha (2001) Central Limit Theorem (CLT) of a linear and quadratic form. Assumption 2.3 is standard in the literature. Assumption 2.4 is exploited in Kelejian and Prucha (1999) to limit the spatial dependence among the units. It rules out any spatial unit root case. As Lee (2004) pointed out, it plays an important role in the derivation of asymptotic properties of the estimators for spatial econometric models. It guarantees that the variance of Y_a and Y_b is bounded as n goes to infinity. Observe that this assumption is also crucial for the identification of the heterogeneous network model, as shown in Proposition 1.

Assumption 4 is a sufficient condition for identification of the social network model. For assumption 4 to hold, $E(Z_a)$ must be full column rank for large enough n_a.

Assumption 2.5. *$F_a = \lim_{n\to\infty} \frac{1}{n} f_a' f_a$ is finite and a full rank matrix, $F_b = \lim_{n\to\infty} \frac{1}{n} f_b' f_b$ is finite and a full rank matrix.*

Since the variance of the structural error is $var(v_a)$ and the concentration parameter (which measures the instrument's strength) is $f_a' f_a / var(v_a)$, this assumption implies that the concentration parameter grows at the same rate as the sample size. Such a rate is assumed in Bekker (1994). Hence, we assume that the instruments are stronger than assumed in the weak-instrument literature.[44] For the sake of brevity we focus on equation (2.2), and we imply the same argument holds for equation (2.3).

Assumption 2.6. *There exists a $K\times(k+2)$ matrix Θ_K such that $\frac{1}{n}\|E(Z_a) - H_K\Theta_K\|^2 \to 0$ as $n, K \to \infty$.*

Following Lee and Liu(2010), assumption 5 requires that the (infeasible) best IV matrix can be well approximated by a certain linear combination of the feasible IV matrix H_K as the number of instruments increases with the sample size. Once we assume this, we can deal with the approximation of S_a and S_{ab}. We have to approximate this matrix since we cannot use it as matrix of instruments because it is formed by unknown parameters. If H_K has the following structure then assumption 5 holds and we can obtain efficiency under certain conditions.

Proposition 5. If $\|\phi_a G_a\|_\infty + \|\phi_{ab}\phi_{ba} C_a\|_\infty < 1$, let us define $H_K^{(p)} = (H_K^{a(p)}, H_K^{b(p)}, X_a^*, G_{ab}X_b)$ where

$$H_K^{a(p)} = (G_a(G_a, (G_{ab}(G_b, \ldots, G_b^{p+1})G_{ba}), \ldots, (G_a(G_a, (G_{ab}(G_b, \ldots, G_b^{p+1})G_{ba})^{p+1}(E(A), G_{ab}J_b B),$$

[44] See Staiger and Stock (1997) or Baltagi et al. (2012) for a panel data version of weak-instrument asymptotics. Another interesting extension could be to derive the estimator's asymptotic properties under many weak instruments. In doing so, we are allowing the rate of concentration parameter to be different than the rate of the sample size. Consequently, we can compare it with the rate in which K increases. See for example, Chao and Swanson (2005)

$$H_K^{b(p)} = (G_{ab}(G_b, (G_{ba}(G_a, \ldots, G_a^{p+1})G_{ab}), \ldots, (G_b(G_b, (G_{ba}(G_a, \ldots, G_a^{p+1})G_{ab})^{p+1}(E(B), G_{ba}J_aA),$$

where p is an increasing integer valued function of K, there exists a $K \times (k+2)$ matrix $\Theta_K^{(p)}$ such that $\|f_a - H_K^{(p)}\Theta_K^{(p)}\|_\infty \to 0 \; as \; n, K \to \infty$.

Therefore, the 2SLS estimator can be asymptotically efficient when we use an increasing number of instruments.

Appendix B: Some Useful Lemmas

Lemma 2.7. *Recall that* $Z_a = f_a + v_a$. *Let* $e_f = \frac{1}{n} f_a'(I - P_K) f_a$. *As* $K/n = O(1)$: *(i)* $Tr(e_f) = o(1)$ *(ii)* $v_a' P_K v_a = O_p(K)$. *(iii)* $f_a' P_K v_a = O_p(\sqrt{nK})$. *(iv)* $e_f = O(tr(e_f))$. *(v)* $\frac{1}{n} Z_a' P_K Z_a = \frac{1}{n} f_a' f_a - e_f + \frac{1}{n} f_a' P_K v_a + \frac{1}{n} v_a' P_K f_a + \frac{1}{n} v_a' P_K v_a = O_p(1)$.

Proof. (i) See lemma B.3 (i) in Lee and Liu (2010).

(ii) Let us write $v_a' P_K v_a = \epsilon_a' S_a' P_K S_a \epsilon_a + \epsilon_b' J_b' G_{ba}' S_{ab}' P_K S_{ab} G_{ba} J_b \epsilon_b$. Let us focus on the first term of the sum, since $E|\epsilon_a' S_a' P_K S_a \epsilon_a| = E[tr(|\epsilon_a' S_a' P_K S_a \epsilon_a|)] = \sigma_a^2 tr(|P_K S_a S_a' P_K|) = O(K)$ by lemma B.2 (ii) Lee and Liu (2010), then by Markov's Inequality $Pr(|\epsilon_a S_a' P_K S_a \epsilon_a|) \geq \alpha) \leq \frac{E(|\epsilon_a S_a' P_K S_a \epsilon_a|)}{\alpha} = O_p(K)$.
For the second part of the sum, given also that $S_{ab} G_{ba} J_b = T_{ab}$ where T_{ab} is UB, we can apply the same proof and obtain the same order of probability. We then have O(f(x))+ O(f(x))=O(f(x)).

(iii) For each j we have by Cauchy-Schwarz inequality

$$|e_j' f_a' P_K v_a| \leq \sqrt{e_j' f_a' f_a e_j} \sqrt{\epsilon_a S_a' P_K S_a \epsilon_a} = O(\sqrt{n}) O_p(\sqrt{K}) = O_p(\sqrt{nK}).$$

(iv) By lemma A.3 (ii) in Donald and Newey (2001).

(v) $\frac{1}{n} Z_a' P_K Z_a = \frac{1}{n} f_a' f_a - O(tr(e_f)) + O_p(K/n) + O_p(\sqrt{K/n}) = O_p(1)$. □

Lemma 2.8. *Recall that* $Z_a = f_a + v_a$, *let* $P_K S_a = \Psi_a$ *and* $P_K T_{ba} = \Xi_{ba}$. *As* $K/n = O(1)$: *(i)* $E(v_a' P_K \epsilon_a) = \sigma_a^2 [e_1, e_2][tr(\Psi_a), \hat{\phi}_{ba}(\Xi_{ba})]'$. *(ii)*$E(v_a' P_K \epsilon_a \epsilon_a' P_K v_a) = \sigma_a^4 tr^2([(\Psi_a), (\Xi_{ba})]) + O(K)$. *(iii)* $[Z_a' P_K \epsilon_a - [\sigma_a^2 [e_1, e_2][tr(\Psi_a), \hat{\phi}_{ba} tr(\Xi_{ba})]' / \sqrt{(n)}] = f_a' \epsilon_a / \sqrt{n} + O_p(\sqrt{K/n}) + O_p(\sqrt{tr(e_f)}) = f_a' \epsilon_a / \sqrt{n} + O_p(1)$.

Proof. (i) $E(v_a' P_K \epsilon_a) = [e_1, e_2] E([(\phi_{ab} S_a G_{ab} J_b \epsilon_b + S_a \epsilon_a), (\phi_{ba} S_{ab} G_{ba} J_a \epsilon_a + S_{ab} \epsilon_b)]') P_K \epsilon_a = [e_1, e_2][E(\epsilon_a' S_a' P_K \epsilon_a), E(\phi_{ba} \epsilon_a' J_a' G_{ba}' S_{ab}' P_K \epsilon_a)]' = \sigma_a^2 [e_1, e_2][tr(\Psi_a), \hat{\phi}_{ba} tr(\Xi_{ba})]'$.

(ii) By lemma A.2 in Lee (2001),

$$E(v_a' P_K \epsilon_a \epsilon_a' P_K v_a) = E([\epsilon_a'(\Psi_a) \epsilon_a \epsilon_a'(\Psi_a) \epsilon_a, \epsilon_a' \Xi_{ba} \epsilon_a \epsilon_a' \Xi_{ba} \epsilon_a]) =$$

$$(\mu_4^a - 3\sigma_a^4)\sum_i [(\Psi_a), \phi_{ba}(\Xi_{ba})]_{ii}^2 + \sigma_0^4[[tr^2(\Psi_a), \phi_{ba}tr(\Xi_{ba})]+$$

$$tr([(\Psi_a), \phi_{ba}(\Xi_{ba})]'[(\Psi_a), \phi_{ba}(\Xi_{ba})] + tr([(\Psi_a)^2, (\Xi_{ba})^2])]$$

$$= \sigma_a^4 tr^2([(\Psi_a), (\Xi_{ba})]) + O(K),$$

where the last equality holds by Lemma B.2 (ii) in Lee and Liu (2010).

(iii) Since $Z_a' P_K \epsilon = f_a \epsilon_a - f_a'(I - P_K)\epsilon + v_a P_K \epsilon$, then

$$(Z_a' P_K \epsilon_a - \sigma_a^2[e_1, e_2][tr(\Psi_a), \hat{\phi}_{ba} tr(\Xi_{ba})])'/\sqrt{(n)} =$$

$$f_a'\epsilon_a/\sqrt{n} - f_a'(I - P_a)\epsilon/\sqrt{n} + [v_a' P_K \epsilon_a - \sigma_a^2[e_1, e_2][tr(\Psi_a), \hat{\phi}_{ba} tr(\Xi_{ba})]']/\sqrt{n}.$$

By Lemma 2.7 above and by Lemma B.2 (ii) in Lee and Liu (2010) $\sqrt{n} f_a'(I - P_a)\epsilon_a = O_p(\sqrt{(Tr(e_f))})$. By Lemma 2 *(i)*, *(ii)* and Markov's inequality for variance we have $\frac{1}{\sqrt{n}}[v_a' P_K \epsilon_a - \sigma_a^2[e_1, e_2][tr(\Psi_a), \hat{\phi}_{ba}(\Xi_{ba})]' = O_p(\sqrt{K/n})$. □

Appendix C: Proofs

Proof of proposition 1. We need to prove that

$$E(Z_a) = (E(G_a y_a), E(G_{ab} y_b), X_a, G_a X_a, G_{ab} X_b)$$

is full column rank. This means that if $E(G_a y_a)d_1 + E(G_{ab} y_b)d_2 + X_a d_3 + G_a X_a d_4 + G_{ab} X_b d_5 = 0$ then $d_1 = d_2 = d_3 = d_4 = d_5 = 0$, where d_1, d_2, d_3, d_4, d_5 are parameters.

By inserting the definitions of $E(G_a y_a)$ and $E(G_{ab} y_b)$ we have:

$$\begin{aligned} G_a(M_a(\phi_{ab} G_{ab} J_b E(B)\delta_b + E(A)\delta_a)d_1 \quad & + G_{ab}(M_b(\phi_{ba} G_{ba} J_a E(A)\delta_b + E(B)\delta_a)d_2 \\ + X_a d_3 + G_a X_a d_4 + G_{ab} X_b d_5 = 0. & \end{aligned}$$

More explicitly,

$$\begin{aligned} G_a(M_a(\phi_{ab} G_{ab} J_b(X_b\beta_b + G X_b \gamma_b + G_{ba} X_a \gamma_{ba}) + X_a\beta_a + G X_a \gamma_a + G_{ab} X_b \gamma_{ab})d_1 & \\ + G_{ab}(M_b(\phi_{ba} G_{ba} J_a(X_a\beta_a + G_a X_a \gamma_a + G_{ab} X_b \gamma_{ab}) + X_b\beta_b + G_b X_b \gamma_b + G_{ba} X_a \gamma_{ba})d_2 & \quad (2.24) \\ + X_a d_3 + G_a X_a d_4 + G_{ab} X_b d_5 = 0 & \quad . \end{aligned}$$

Let us assume that J_a, J_b, M_a and M_b are invertible and thus

$J_a = (\phi_a G_a)^{-1} = \sum_{k=0}^{\infty} (\phi_a G_a)^k,$

$J_b = (\phi_b G_b)^{-1} = \sum_{k=0}^{\infty} (\phi_b G_b)^k,$

$M_a = (\phi_a G_a + \phi_{ab}\phi_{ba} G_{ab} J_b G_{ba})^{-1} = \sum_{j=0}^{\infty} (\phi_a G_a + \phi_{ab}\phi_{ba} G_{ab} \sum_{k=0}^{\infty} (\phi_b G_b)^k G_{ba})^j,$

$M_b = (\phi_b G_b + \phi_{ba}\phi_{ab} G_{ba} J_a G_{ab})^{-1} = \sum_{j=0}^{\infty} (\phi_b G_b + \phi_{ba}\phi_{ab} G_{ba} \sum_{k=0}^{\infty} (\phi_a G_a)^k G_{ab})^j.$

Going back to equation (2.25), we obtain

$$
\begin{aligned}
G_a(\sum_{j=0}^{\infty}(\phi_a G_a + \phi_{ab}\phi_{ba}G_{ab}\sum_{j=0}^{\infty}(\phi_b^j G_b^j)G_{ba})^j(\phi_{ab}G_{ab}\sum_{k=0}^{\infty}(\phi_b G_b)^k(X_b\beta_b + G_b X_b\gamma_b + G_{ba}X_a\gamma_{ba})) \\
+X_a\beta_a + G_a X_a\gamma_a + G_{ab}X_b\gamma_{ab})d_1 \\
+G_{ab}(\sum_{j=0}^{\infty}(\phi_b G_b + \phi_{ba}\phi_{ab}G_{ba}\sum_{k=0}^{\infty}(\phi_a G_a)^k G_{ab})^j(\phi_{ba}G_{ba}\sum_{k=0}^{\infty}(\phi_a G_a)^k(X_a\beta_a + G_a X_a\gamma_a + G_{ab}X_b\gamma_{ab}))+ \\
X_b\beta_b + G_b X_b\gamma_b + G_{ba}X_a\gamma_{ba})d_2 \qquad (2.25) \\
+X_a d_3 + G_a X_a d_4 + G_{ab}X_b d_5 \quad = 0 \quad .
\end{aligned}
$$

The left side of the previous equation is the sum of products of the matrices G_a, G_{ba}, G_{ab} and G_b times X_a or X_b weighted by different parameters.[45]

Let us define $J = (k, p, m)$ and $C(c(1) \in A, \cdot, ..., c(l) \in B, J)$. C is a *set of paths*, hereafter called a *chain*,[46] of length l which starts from A and ends at B, having k links from a type B node to another type B node, p links from a type A node to another type A node, and m links between nodes of different types. The concept of *chain* is particularly useful in our context. Indeed, the product of adjacency matrices contains the same information of a chain. For instance $G_a \equiv C(c(1) \in A, c(2) \in A, k = 0, p = 1, m = 0)$ and $G_a G_{ab} \equiv C(c(1) \in A, c(2) \in A, c(3) \in B, k = 0, p = 1, m = 1)$. A similar characterization can be written for all combinations (products) of adjacency matrices considered in equation (2.25).

Taking advantage of this notation, the system in equation (2.25) can be characterized by the following two matrices

[45] The matrices sequence is multiplied by X_a or X_b depending on the last interaction matrix. For instance $G_a^2 G_{ab}$ is multiplied by X_b while $G_b G_{ba}$ is multiplied by X_a.

[46] In this notation a *chain* includes all possible *paths* that have common features. For instance, all of *paths* starting from A and arriving to B are in the same *chain*.

$$C=\left[\frac{C_a}{C_{ab}}\right]=\begin{bmatrix} \\ I_a \\ G_a \\ G_a^2 \\ . \\ G_a^k \\ . \\ G_{ab}G_{ba} \\ . \\ G_{ab}G_bG_{ba} \\ . \\ G_{ab}G_b^kG_{ba} \\ . \\ G_aG_{ab}G_bG_{ba} \\ . \\ G_a^lG_{ab}G_b^kG_{ba} \\ G_aG_{ab}G_bG_{ba} \\ . \\ G_a^kG_{ab}G_b^lG_{ba} \\ . \\ C(c(1)\equiv A,\cdot,c(l)\equiv A,J) \\ . \\ \hline G_{ab} \\ G_aG_{ab} \\ . \\ G_a^kG_{ab} \\ . \\ G_{ab}G_b \\ . \\ G_{ab}G_b^k \\ . \\ G_aG_{ab}G_bG_{ba}G_{ab}G_{ba} \\ . \\ G_a^lG_{ab}G_b^kG_{ba}G_{ab}G_{ba} \\ . \\ G_{ab}G_bG_{ba}G_aG_{ab} \\ . \\ G_{ab}G_b^kG_{ba}G_a^lG_{ab} \\ . \\ C(c(1)\equiv A,\cdot,c(l)\equiv B,J) \\ . \end{bmatrix}, \Theta=\left[\frac{\Theta_a}{\Theta_{ab}}\right]=\left[\begin{array}{ccccc} d_1 & d_2 & d_5 & d_4 & d_3 \\ \hline 0 & 0 & 0 & 0 & 1 \\ \beta_a & 0 & 0 & 1 & 0 \\ \beta_a\phi_a+\gamma_a & 0 & 0 & 0 & 0 \\ . & . & . & . & . \\ (\beta_a\phi_a+\gamma_a)\phi_a^{k-1} & 0 & 0 & 0 & 0 \\ . & . & . & . & . \\ 0 & \gamma_{ba} & 0 & 0 & 0 \\ . & . & . & . & . \\ 0 & \gamma_{ba}\phi_b & 0 & 0 & 0 \\ . & . & . & . & . \\ 0 & \gamma_{ba}\phi_b^k & 0 & 0 & 0 \\ . & . & . & . & . \\ \beta_a\phi_{ba}\phi_{ab}\phi_b\phi_a & 0 & 0 & 0 & 0 \\ . & . & . & . & . \\ \beta_a\phi_{ba}\phi_{ab}\phi_b^k\phi_a^l & 0 & 0 & 0 & 0 \\ \beta_b\phi_{ba}\phi_{ab}\phi_b^k & 0 & 0 & 0 & 0 \\ . & . & . & . & . \\ \beta_b\phi_{ba}\phi_{ab}\phi_b^k & 0 & 0 & 0 & 0 \\ . & . & . & . & . \\ f(\theta,J)g(\Phi,J) & f(\theta,J)g(\Phi,J) & 0 & 0 & 0 \\ . & . & . & . & . \\ \hline 0 & \beta_b & 1 & 0 & 0 \\ \gamma_{ab} & 0 & 0 & 0 & 0 \\ . & . & . & . & . \\ \gamma_{ab}\phi_a^{k-1} & 0 & 0 & 0 & 0 \\ . & . & . & . & . \\ 0 & (\gamma_b+\beta_b\phi_b) & 0 & 0 & 0 \\ . & . & . & . & . \\ 0 & (\gamma_b+\beta_b\phi_b)\phi_b^{k-1} & 0 & 0 & 0 \\ . & . & . & . & . \\ \gamma_{ba}\phi_{ba}\phi_{ab}\phi_b & 0 & 0 & 0 & 0 \\ . & . & . & . & . \\ \gamma_{ba}\phi_a^{l-1}\phi_{ba}\phi_{ab}\phi_b^k & 0 & 0 & 0 & 0 \\ . & . & . & . & . \\ 0 & \beta_b\phi_{ba}\phi_{ab}\phi_a\phi_b & 0 & 0 & 0 \\ . & . & . & . & . \\ 0 & \beta_b\phi_{ba}\phi_{ab}\phi_a^l\phi_b^k & 0 & 0 & 0 \\ . & . & . & . & . \\ f(\theta,J)g(\Phi,J) & f(\theta,J)g(\Phi,J) & 0 & 0 & 0 \\ . & . & . & . & . \end{array}\right],$$

where C represents different products of G_a, G_{ba}, G_{ab} and G_b (*chains*) appearing in the left side of equation (2.25) and Θ collects the relative coefficients. Θ has five columns which distinguish the elements that are multiplied by d_1, d_2, d_3, d_4 or d_5.

The lower panel represents *chains* starting from A and arriving to B (labeled as C_{ab}), while the upper panel collects *chains* starting from A and coming back to A (labeled as C_a). The generic element of Θ is defined by the following objects, $\theta = (\theta_a, \theta_b)$, where $\theta_a = (\beta_a, \gamma_a, \gamma_{ab})$ and $\theta_b = (\beta_b, \gamma_b, \gamma_{ba})$; $\Phi = (\phi_{ba}, \phi_{ab}, \phi_b, \phi_a)$ and $f(\theta, J) = \beta_a I_{J,\beta_a}(\beta_a) + \gamma_a I_{J,\gamma_a}(\gamma_a) + \gamma_{ab} I_{J,\gamma_{ab}}(\gamma_{ab}) + \beta_b I_{J,\beta_b}(\beta_b) + \gamma_b I_{J,\gamma_b}(\gamma_b) + \gamma_{ba} I_{J,\gamma_{ba}}(\gamma_{ba})$ is a set of indicator functions that take value one if the argument appears in the corresponding chain and zero otherwise. The function $g(\Phi, J) = \prod_k \phi_b^k \prod_l \phi_a^l \prod_m (\phi_{ba}\phi_{ab})^m$ keeps track of the number of times the relative chain passes from one type of node to another scaled by the respective interaction parameters. Observe that $(d_1, d_2, d_3, d_4, d_5)'\Theta' H_\infty = 0$ is equal to the condition $E(G_a y_a)d_1 + E(G_{ab} y_b)d_2 + X_a d_3 + G_a X_a d_4 + G_{ab} X_b d_5 = 0$. The elements of H_∞ are equal to the elements of C multiplied by X_a or X_b depending on the last interaction matrix.[47]

From C and Θ one can argue that the model is identified in the cases listed in Proposition 1.

Let us focus on case (1). For $E(Z_a)$ to have full rank, it suffices that Θ has full rank. This means that we need the linear independence of at least five *chains* (rows of C), translating to the linear independence of I_a, G_a, G_a^2, G_{ab} and $G_{ab}G_b$. [48] The corresponding five rows of Θ are thus linear independent. Additionally we need to have five linear independent columns of Θ, so having $\beta_a\phi_a + \gamma_a \neq 0$ and $\beta_b\phi_b + \gamma_b \neq 0$ suffices to reach the full rank condition for Θ and consequently

[47] Note that H_∞ is the IV matrix considered in Section 2.4, which is approximated by H_K in the feasible 2SLS estimation.

[48] Note that here we need at least three chains from C_a and two from C_{ab} because we are considering the outcome equation for type A nodes, i.e. the staring point of chains is always a type A node.

$E(Z_a)$. The same argument applies for case (2).

□

Relationship with *chains* and *trees*. In the proof of Proposition 1 we have established the equivalence between sequences of products of adjacency matrices and the concept of *chains*. In order to provide a better intuition behind the multiple sufficient conditions argument note that, according to the proof of Proposition 1 notation, a set of chains with a certain length p can be divided in g^{p+1} number of chains, where g is the number of node types. For instance, chains of length 1 can be classified in four categories when nodes are split into two types. following proof notation we can define $C(1) \equiv C(c(1) \equiv a, c(2) \in A, 0, 1, 0) \cup C(c(1) \in B, c(2) \in B, 1, 0, 0) \cup C(c(1) \in A, c(2) \in B, 0, 0, 1) \cup C(c(1) \in B, c(2) \equiv a, 0, 0, 1)$ (e.g. $G_a \cup G_b \cup G_{ab} \cup G_{ba} = G$).

We can see this system of chains as a tree, more specifically as a *Tree-indexed Markov chain*. A tree is a graph with a distinguished vertex $x_0 \in g$ (here a type A node, the starting point) and the degree of each vertex is at least two (in our case the number of types, g). Its structure is basically determined by a countable set of states (in our case the number of types, g) characterized by a transition probability ($\{p(x, y)|x, y \in g\}$ in our case).[49]

Let $T_a := \cup_{l,J} C(c(1) \in A, \cdot, c(l), J)$ (Figure 2.3), it is simply the collection of all possible chains of all possible lengths starting in a type A node. For identification purposes, we simply need that G_a, G_b, G_{ab} and G_{ba} are not empty (and not full).[50] In words, it means that there are no reasons why two randomly drawn nodes cannot be connected for each combination type (or that each node

[49]Given that here we are not interested in determining the transition probability law of a chain, even if it is simple to estimate and is basically the link formation probability considered for all of the possible combinations of nodes' type. Benjamini and Peres (1994) give a detailed discussion on *Tree-indexed Markov chain.*

[50]It is equivalent to say that the probability $0 < P(g_{ij} = 1) < 1$, $i \in A, B$ and $j \in A, B$. Note that transition probability can be derived from G_a, G_b, G_{ab} and G_{ba}. Here we are simply excluding the classical linear in mean framework (when the matrices are complete) and the case in which there are no connections (when the matrices are empty).

Figure 2.3: Tree of Chains when type A and B nodes are considered (or have connections)

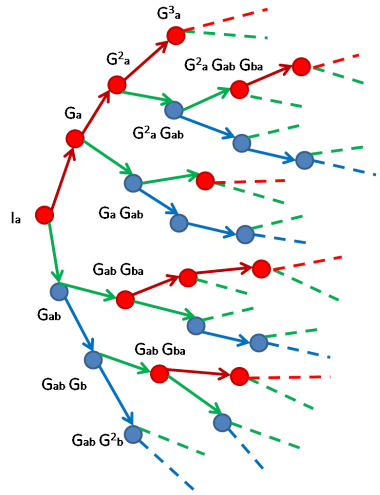
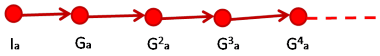


Figure 2.4: Non-tree of chains when only type A nodes are considered (or have connections)



is connected with all of other nodes).[51]

An interesting feature of this framework is that, as case (2) tells us, even if I_a, G_a, and G_a^2 are linearly dependent, we can still identify ϕ_a and the other parameters relying on linear independence of chains passing through type B nodes.[52] In other words, we can identify the parameters because of the multiple branches of the tree (see Figure 2.3).[53]

[51] From a Markov Chain perspective again, a more restrictive condition consists in assuming that the underlying Markov Chain is irreducible and aperiodic. This means that type A are connected with type B or type A with the same probability (and the same holds for type B). Thus, in this case tree branches with the same length have the same probability of being observed. The aperiodicity and irreducibility are not necessary for the identification condition to hold, but of course are sufficient.

[52] Holding condition (2) instead of (1). We basically take advantage of linear independence of I_a, G_a and $G_{ab}G_{ba}$ instead of G_a^2

[53] The additional parameter restrictions (conditions (1b, 2a or 2b) in Proposition 1) are basically due to an additional vector in the full rank condition (i.e. $E(G_{ab}y_b)$).

Comparison with the identification conditions for *homogeneous* models. Let us conclude this discussion by further highlighting the connection between identification in a single group model and in a multiple group one. Let us reproduce a single group model by considering only type A nodes. The model is

$$y_a = \phi_a G_a y_a + \beta_a X_a + \gamma_a G_a X_a + \epsilon_a. \tag{2.26}$$

In order to obtain identification we want $(E(G_a y_a), X_a, G_a X_a)$ to have full rank. Given that $E(G_a y_a) = G_a(I - \phi_a G_a)^{-1}(\beta_a X_a + \gamma_a G_a X_a) = G_a \sum_{j=0}^{\infty}(\phi_a G_a)^j(\beta_a X_a + \gamma_a G_a X_a)$, the matrices used in the proof of Proposition 1 can be written in the following way

$$C = \begin{bmatrix} C_a \end{bmatrix} = \begin{bmatrix} I_a \\ G_a \\ G_a^2 \\ . \\ G_a^k \\ . \end{bmatrix}, \Theta = \begin{bmatrix} \Theta_a \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ \beta_a & 1 & 0 \\ \beta_a\phi_a + \gamma_a & 0 & 0 \\ . & . & . \\ (\beta_a\phi_a + \gamma_a)\phi_a^{k-1} & 0 & 0 \\ . & . & . \\ & & \end{bmatrix}.$$

As before, these two matrices respectively represent the chains and their coefficients. According to the proof of Proposition 1, the full rank condition for $(E(G_a y_a), X_a, G_a X_a)$ depends on C and Θ. From C and Θ one can argue that the model is identified if (see proof of Proposition 1 for details)

1. $\beta_a\phi_a + \gamma_a \neq 0$,

2. I_a, G_a and G_a^2 are linear independent.

These are exactly the conditions of Proposition 1 in Bramoullé et al. (2009). Note that if I_a, G_a and G_a^2 are linearly dependent, then G_a^k is also linearly dependent $\forall k$. Given that here we cannot differentiate nodes, we have $I_a \equiv C(0)$, $G_a \equiv C(1)$, and $G_a^2 \equiv C(2)$, where $C(k)$ represent the set of chains with length k. In terms of chains it means that $C(k) \equiv \{C(k-1), C(1)\} \equiv \{C(k-2), C(2)\} \equiv \cdots \equiv \{C(2), C(k-2)\} \equiv \{C(1), C(k-1)\}$. In words it means that each chain's set can be represented by at least two sets of chains. So each G_a^k can be represented by the product of two matrices, $G_a G_a^{k-1}$, $G_a^2 G_a^{k-2}$, and so on. This is the connection to the linear independence of I_a, G_a, and G_a^2 as a condition for identification. In this case, a length l set of chains cannot be separated by node type, and thus T_a is composed only of one chain (Figure 2.4) instead of multiple chains (Figure 2.3).[54] Therefore, we need I_a, G_a, and G_a^2 to be linearly independent in order to have at least three independent chains in C and consequently identify the model's parameters satisfying the restriction $\beta_a \phi_a + \gamma_a \neq 0$.

Proof of proposition 2. By the classical expansion the estimator is

$\sqrt{n}(\hat{\mu} - \mu_0) = \frac{1}{n}(Z_a' P_K Z_a)^{-1} \frac{1}{\sqrt{n}} Z_a' P_K \epsilon_a$. As $Z_a = f_a + v_a$, by Lemma 2.7 *(v)*, we have $\frac{1}{n}(Z_a' P_K Z_a) = F_a + o_p(1)$. By Lemma 2.8 *(iii)*

$$[Z_a' P_K \epsilon_a - \sigma_a^2 [e_1, e_2][tr(\Psi_a), \hat{\phi}_{ba}(\Xi_{ba})]']/\sqrt{(n)} = f_a' \epsilon_a / \sqrt{n} + o_p(1) \xrightarrow{d} N(0, \sigma_a^2 F_a)$$

by CLT.

Hence, the proposition is derived by Slutzky theorem

$$\frac{1}{n}(Z_a' P_K Z_a)^{-1} \frac{1}{\sqrt{n}} Z_a' P_K \epsilon \xrightarrow{d} F_a^{-1} \cdot N(0, \sigma_a^2 F_a) = N(0, \sigma_a^2 F_a^{-1}).$$

□

Proof of proposition 3. Given the proof of Proposition 2, it is sufficient to

[54] Borrowing from Markov chains vocabulary again, this is because the state that characterizes the chain is only one (A).

show that

$$\hat{\sigma}_a^2[e_1, e_2][tr(\Psi_a), \hat{\phi}_{ba}(\Xi_{ba})]'/\sqrt{n} = o_p(1).$$

If we fix C_a, then by Lemma C.12 in Lee and Liu (2008) $\hat{M}_a - M_a = \hat{M}_a(\hat{\phi}_a - \phi_a)G_a + \hat{M}_a((\hat{\phi}_{ab}\hat{\phi}_{ba} - \phi_{ab}\phi_{ba})C_a)$. So we can write, $tr(\hat{\Psi}_a) - tr(\Psi_a) = tr(P_K(S_a(\hat{\phi}) - S_a) = tr(P_K(G_a(\hat{M}_a - M_a) = tr(P_K(G_a(\hat{M}_a(\hat{\phi}_a - \phi_a)G_a + \hat{M}_a((\hat{\phi}_{ab}\hat{\phi}_{ba} - \phi_{ab}\phi_{ba})C_a) = (\hat{\phi}_a - \phi_a)tr(P_K(G_a(\hat{M}_a G_a) + (\hat{\phi}_{ba}\hat{\phi}_{ba} - \phi_{ab}\phi_{ba})tr(P_K(G_a(\hat{M}_a C_a))$.

Since the product of UB matrices is still UB (Kelejian and Prucha 1998), using the lemma B.2 (ii) in Lee and Liu (2010) and the initial $\sqrt{n}$ consistency assumption, we obtain

$\sqrt{n}(\hat{\phi}_a - \phi_a)tr(P_K(G_a(\hat{M}_a G_a) + \sqrt{n}(\hat{\phi}_{ba}\hat{\phi}_{ba} - \phi_{ab}\phi_{ba})tr(P_K(G_a(\hat{M}_a C_a))/n = o_p(1)O(K/n) + o_p(1)O(K/n) = o_P(K/n)$.

Finally, we have $\sqrt{n}(\hat{\sigma}_a^2 - \sigma_a^2)(tr(\hat{\Psi}_a) - tr(\Psi_a))/n = o_p(1)o_p(K/n) = o_p(K/n) = o_P(1)$ as $K/n \to 0$. The same procedure can be applied for $\sqrt{n}(\hat{\sigma}_b^2 - \sigma_b^2)(tr(\hat{\Xi}_a) - tr(\Xi_a))/n$ and for the second element of the stacked vector v. □

Proof of proposition 4. Let p be a finite integer. Let us define the number of instruments equal to

$$K = \sum_{p=1}^{P} g^p + o(1),$$

so that we have $\sum_{n=1}^{P} g^p = O(g^p)$. Since we assume $K/n \to 0$, we have $\sum_{p=1}^{P} g^p = o(n)$ by assumption. This implies that $g^p = o(n)$. It follows that $g = o(n^{1/p})$. □

Proof of proposition 5. We prove this proposition for $H_K^{a(p)}$. The same applies for $H_K^{b(p)}$. Let $\Theta_K^{(p)}$ be the matrix of true parameters derived from the *p-order* expansion of Θ (see Section 2.3). If $\sup\|\phi_a G_a\|_\infty + \sup\|\phi_{ab}\phi_{ba}C_a\|_\infty < 1$, then

$$H_K^{(p)}\Theta_K^{(p)} = G_a \sum_{j=0}^{p} (\phi_a G_a + \phi_{ab}\phi_{ba}C_a^k)^j(\phi_{ab}G_{ab}J_b E(B)\delta_b + E(A)\delta_a).$$

It follows that $\|f_a - H_K^{(p)}\Theta_K^{(p)}\|_\infty = \|(\phi_a G_a + \phi_{ab}\phi_{ba}C_a^k)^{p+1}S_a(\phi_{ab}G_{ab}J_bE(B)\delta_b + E(A)\delta_a)\|_\infty \leq \|(\phi_a G_a + \phi_{ab}\phi_{ba}C_a^k)^{p+1}\|_\infty \|S_a\|_\infty \|\phi_{ab}G_{ab}J_bE(B)\delta_b + E(A)\delta_a\|_\infty = o(1)\ as\ p \to \infty.$ □

Appendix D: Tables and Figures

Table 2.1: Monte Carlo Simulation: 1000 obs., 1000 replications

	(1)	(2)	(3)
10 max connections	2SLS finite IVs	2SLS large IVs	2SLS bias-corrected
$\phi_a = 0.1$	0.100(0.032)	0.100(0.026)	0.099 (0.027)
$\phi_{ab} = 0.2$	0.201(0.031)	0.201(0.026)	0.197(0.069)
$\beta_a = 0.5$	0.501(0.047)	0.501(0.047)	0.501(0.048)
$\gamma_a = 0.5$	0.503(0.081)	0.502(0.078)	0.503(0.076)
$\gamma_{ab} = 0.5$	0.496(0.079)	0.496(0.075)	0.500(0.097)
20 max connections	2SLS finite IVs	2SLS large IVs	2SLS bias-corrected
$\phi_a = 0.1$	0.098(0.025)	0.098(0.020)	0.100(0.020)
$\phi_{ab} = 0.2$	0.197(0.023)	0.195(0.019)	0.200(0.019)
$\beta_a = 0.5$	0.501(0.048)	0.501(0.048)	0.501(0.048)
$\gamma_a = 0.5$	0.506(0.096)	0.506(0.093)	0.503(0.093)
$\gamma_{ab} = 0.5$	0.500(0.097)	0.497(0.093)	0.496(0.093)
30 max connections	2SLS finite IVs	2SLS large IVs	2SLS bias-corrected
$\phi_a = 0.1$	0.099(0.020)	0.098(0.016)	0.099(0.016)
$\phi_{ab} = 0.2$	0.198(0.019)	0.195(0.015)	0.199(0.015)
$\beta_a = 0.5$	0.500(0.048)	0.501(0.047)	0.501(0.047)
$\gamma_a = 0.5$	0.506(0.110)	0.507(0.107)	0.505(0.107)
$\gamma_{ab} = 0.5$	0.500(0.112)	0.498(0.109)	0.497(0.109)

Note: y_b is generated with $\phi_b = 0.1$, $\phi_{ba} = 0.2$, $\beta_b = 0.5$, $\gamma_b = 0.5$, $\gamma_{ba} = 0.5$

Table 2.2: Monte Carlo Simulation: 1000 obs., 1000 replications

	(1)	(2)	(3)
20 max connections	2SLS few IVs	2SLS many IVs	2SLS bias-corrected
$\phi_a = 0.1$	0.099(0.040)	0.095(0.032)	0.097(0.032)
$\phi_{ab} = 0.1$	0.101(0.038)	0.101(0.031)	0.099(0.031)
$\beta_a = 0.5$	0.501(0.047)	0.505(0.047)	0.501(0.047)
$\gamma_a = 0.5$	0.504(0.084)	0.506(0.081)	0.506(0.080)
$\gamma_{ab} = 0.5$	0.497(0.083)	0.497(0.078)	0.498(0.078)

Note: y_b is generated with $\phi_b = 0.1$, $\phi_{ba} = 0.1$, $\beta_b = 0.5$, $\gamma_b = 0.5$, $\gamma_{ba} = 0.5$

20 max connections	2SLS few IVs	2SLS many IVs	2SLS bias-corrected
$\phi_a = 0.1$	0.097(0.029)	0.097(0.022)	0.103(0.022)
$\phi_{ab} = 0.3$	0.298(0.027)	0.298(0.021)	0.302(0.021)
$\beta_a = 0.5$	0.501(0.048)	0.501(0.048)	0.501(0.048)
$\gamma_a = 0.5$	0.506(0.077)	0.506(0.075)	0.501(0.075)
$\gamma_{ab} = 0.5$	0.500(0.076)	0.500(0.072)	0.496(0.072)

Note: y_b is generated with $\phi_b = 0.1$, $\phi_{ba} = 0.3$, $\beta_b = 0.5$, $\gamma_b = 0.5$, $\gamma_{ba} = 0.5$

20 max connections	2SLS few IVs	2SLS many IVs	2SLS bias-corrected
$\phi_a = 0.1$	0.099(0.016)	0.097(0.010)	0.091(0.452)
$\phi_{ab} = 0.4$	0.390(0.012)	0.370(0.008)	0.401(0.064)
$\beta_a = 0.5$	0.501(0.048)	0.501(0.048)	0.502(0.079)
$\gamma_a = 0.5$	0.505(0.074)	0.504(0.073)	0.502(0.073)
$\gamma_{ab} = 0.5$	0.498(0.070)	0.478(0.069)	0.498(0.582)

Note: y_b is generated with $\phi_b = 0.2$, $\phi_{ba} = 0.05$, $\beta_b = 0.5$, $\gamma_b = 0.5$, $\gamma_{ba} = 0.5$

Table 2.3: Monte Carlo Simulation: 1000 obs., 1000 replications, 10 max connections, $\phi_b = \phi_{ab} = \phi_{ba} = 0.3$,$\beta_b = 0.5$,$\gamma_b = \gamma_{ab} = \gamma_{ba} = 0.5$

	2SLS/Misspeficied model	2SLS/misspecified IVs	2SLS/correct model-correct IVs
$\phi_a = 0.3$	0.3684 (0.0335)	0.2971 (0.0465)	0.2999 (0.0164)
$\beta_a = 0.5$	0.3856 (0.1804)	0.4865 (0.2233)	0.5097 (0.1489)
$\gamma_a = 0.5$	-0.0016 (0.2315)	0.4929 (0.2102)	0.4963 (0.1765)

Table 2.4: Monte Carlo Simulation: 1000 obs., 1000 replications, 10 max connections

	(1) $\phi_a = \phi_b = \phi_{ba} = \phi_{ab} = 0.1$	(2) $\phi_a = \phi_b = 0.1$,$\phi_{ba} = \phi_{ab} = 0.3$	(3) $\phi_a = 0.1$, $\phi_b = 0.2$, $\phi_{ba} = 0.05$, $\phi_{ab} = 0.4$
ϕ	0.100(0.020)	0.178(0.021)	0.205(0.022)
β	0.500(0.031)	0.499(0.034)	0.499(0.033)
γ	0.446(0.022)	0.445(0.024)	0.442(0.025)

Figure 2.5: Policy experiment: varying ϕ_a and ϕ_{ab}

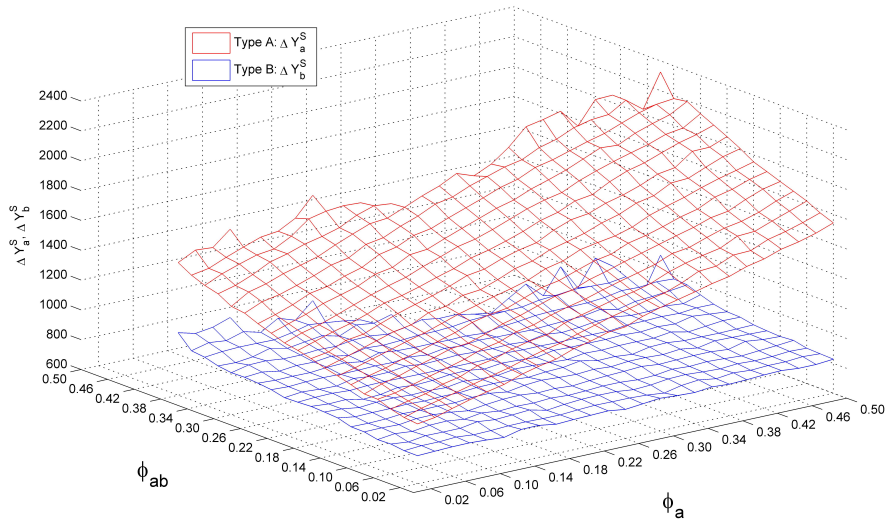


Figure 2.6: Policy experiment: varying ϕ_a and ϕ_{ba}

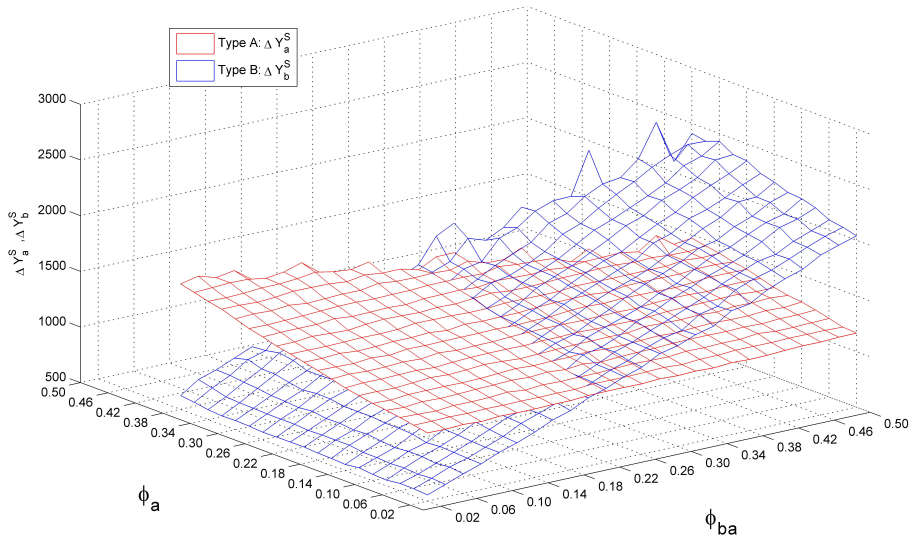


Figure 2.7: Kernel density estimation of empirical distributions of $\Delta y_{i\in A}$ and $\Delta y_{i\in B}$, increasing ϕ_a.

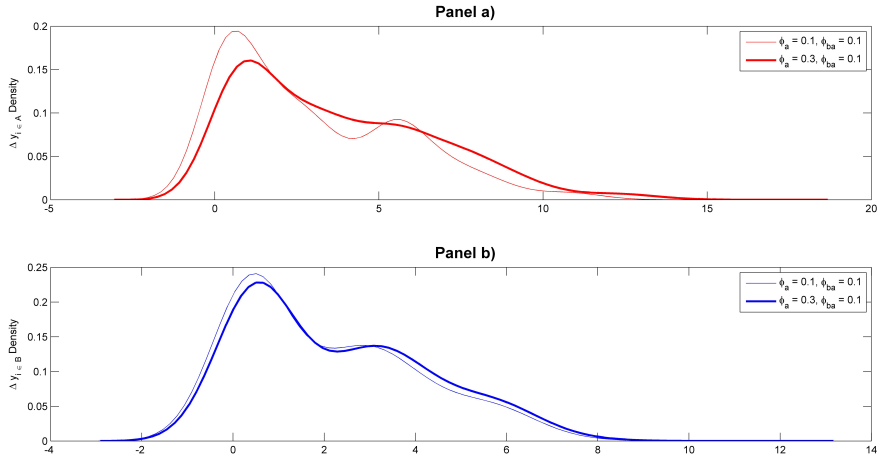
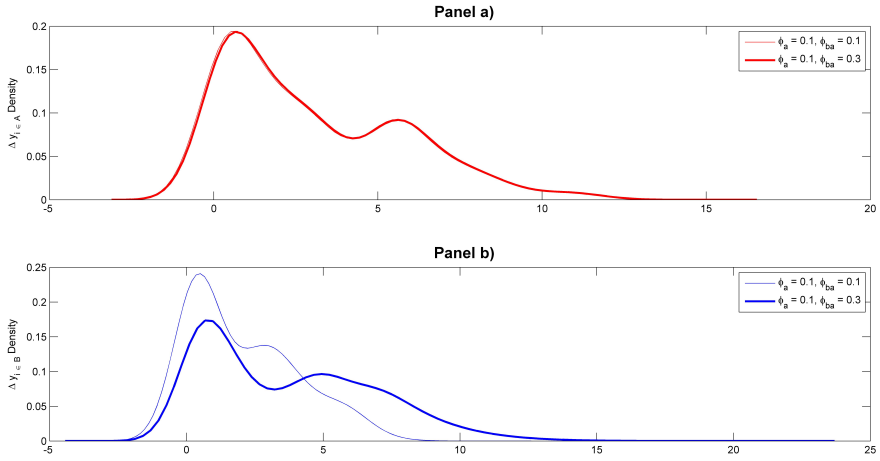


Figure 2.8: Kernel density estimation of empirical distributions of $\Delta y_{i \in A}$ and $\Delta y_{i \in B}$, increasing ϕ_{ba}.



Chapter 3

The Word on Banking: Social Ties, Trust, and the Adoption of Financial Products

Joint work with Eleonora Patacchini.

Another version of this chapter has been published as *EIEF Working Paper No. 14/04*

3.1 Introduction

What factors drive banking decisions? How people choose financial products? A recent study conducted by the Financial Brand in 2011 reveals that, in the previous two years, the percentage of consumers choosing online and offline word-of-mouth (i.e. face-to-face) as the most important driver of banking product and service purchases has increased significantly, whereas the share of those reporting past experience as the crucial factor slightly decreased.[1] The face-to-face channel drives about a third of consumers' checking, savings and mortgage account

[1]The report is based on the Large Purchase Study conducted by S. Radoff Associates in summer 2010 on a nationally representative sample of 1,000 U.S. adults aged 18 and up.

choices. It also explains about one-quarter of credit card brand choices. When looking at the factors influencing banking decisions by age groups, the study reveals that for young people (18-29 years old), face-to-face communication is the most important factor. Its share of almost 50% largely dominates both past experience and online word-of-mouth (both with shares lower than 30%). The low importance of past experience is expected because of the young age of this group, but why face-to-face social contacts are more important than online social contacts is not obvious This may be further puzzling for those who think online/social media has tremendous power to influence a large number of consumers.

Using an unique data set of friendships among a representative sample of United States students, we investigate the role of social interactions for financial decisions during the early adulthood.

A major challenge in the empirical analysis of peer effects with non-experimental data is the ability to distinguish correlations in behaviors of individuals within a group from correlated preferences of people that sort into the same group. Most of the existing studies (see Jackson and Zenou; 2014) resort on the use of network fixed effects to account for sorting into the same network (group formation). The underlying assumption is that conditional on observed individual characteristics and network-level unobserved characteristics, peers within networks are chosen at random. A possible sorting into relationships (peer group formation) along unobservable (individual-level) characteristics remains a major concern.

In this paper, we use Bayesian inferential methods to integrate network formation with the study of behavior over the formed network. We show that such an approach not only tackles endogeneity issues arising from sorting into relationships, but also those related to missing peers and/or missing links between peers.

Our identification strategy hinges on three main features, which are novel to the financial literature. First, the uniqueness of our dataset lies in the fact that

it is based on direct friends' nominations and provides complete information on all nominated friends. This allows us to control not only for individual characteristics but also for peers' characteristics, thus controlling for sorting (into peer groups) along observed characteristics. Second, because we observe individuals over networks we can employ a pseudo-panel data method and control for network fixed effects. As mentioned above, this strategy helps accounting for sorting along *network-level* unobserved dimensions, given that the influence of any factor which is constant across individuals in the same network is washed away. Third, we account for sorting along *individual-level* unobserved factors by modelling jointly network formation and behavior over networks using a bayesian approach. This strategy enables to control for the influence of individual-level unobserved factors that might affect both friends' choice and financial decisions. This is the important methodological innovation used in this paper.

Our analysis uncovers one main novel and important feature. We find that not all social contacts are equally important: only those with a long-lasting relationship (strong ties) influence financial decisions. Moreover, the correlation in agents' behavior only arises among strong ties in cohesive network structures. The length of the relationship does not seem to proxy for its intensity. The richness of our datasets allows us to distinguish between the two effects, finding that it is the length of time spend together that matters the most.

This evidence is consistent with the literature in finance showing an important role of trust in financial decisions (see, most notably Guiso et al.; 2008, 2004).[2] When agents have to decide whether to adopt or not a financial instrument they face a risk and they might value more the information coming from agents they trust. Our analysis thus helps understanding why face-to-face social contacts are more important than online social contacts. Online word-of-mouth can be seen

[2]Butler et al. (2012) highlight financial advice as an important example of trust-based exchange. In the US, 73% of all retail investors consult a financial advisor before purchasing shares (Hung et al.; 2008).

as a less reliable source of information, since the agents spreading the information are not personally known and consequently not necessarily trustworthy.

Financial and payment instruments are fundamental in the economy smooth functioning as a support for money and asset transfers among agents. The adoption of novel and technology-based financial instruments are trust-intensive decisions, people might trust other people when collecting private information about a specific financial product.[3] The role of social interactions is thus at the crux of a full understanding of potential diffusion of technological changes.[4]

There are only a few papers that look at the importance of social interactions in finance.

Hong et al. (2004) find that social households, as defined as those who interact with their neighbors or attend church, are more likely to invest in the stock market than non-social households. They present a model where social investors differ from non-socials in that their net cost of participating in the market is influenced by the choices of their peers.[5] Their model predicts an higher participation rate among social investors than among non-socials, and also that a social multiplier is likely to arise from the correlation between individual and peers' financial decisions. Because of the absence of information on precise social interaction patterns in their data (i.e. about who interacts with whom), their empirical analysis focuses on testing the first model prediction only.[6] Our anal-

[3] Algan and Cahuc (2014) characterize trust as an important driver of economic development, and identify financial markets as one of the main channels through which trust influences economic outcomes of a society. The relationship between *individual* trust and *individual* economic outcomes is investigated by Butler et al. (2010).

[4] Economists have been optimistic that currency will be replaced by technologically more advanced electronic transfers and e-moneys of assorted varieties (see, e.g. Craig; 1999; Drehmann et al.; 2002). The cost of a country's payment system is usually between 2% and 3% of GDP. Since the cost of an electronic payment ranges between one-third to one-half that of a check or paper giro payment (see, e.g. Gresvik; 2009; Humphrey and Berger; 1990), promoting a shift to electronic would reduce this cost. In addition, the use of cash is affected by the extent of illegal activities including the avoidance of taxes (see, e.g. Humphrey et al.; 1996).

[5] Specifically, the cost for any social investor in a given peer group is reduced—relative to the value for an otherwise identical non-social—by an amount that is increasing in the number of others in the peer group that are participating.

[6] They provide evidence consistent with a peer-effects story by finding that the impact of sociability is stronger in states where stock-market participation rates are higher.

ysis complements their findings, as it provides evidence on the existence and the extent of the social multiplier in financial decisions. As Hong et al. (2004) argue, the presence of a social multiplier may help understanding changes in aggregate stock-market participation over time. If the increase of stock market participation in the last decades can be associated with a decrease in participation cost, then social interactions may have had a crucial role by amplifying the cost shock.

Using a high-stakes field experiment conducted with a financial brokerage, Bursztyn et al. (2012) find that both social learning and social utility channels have statistically and economically significant effects on investment decisions. Indeed, a peer's act of purchasing an asset would affect one's own choice because one may acquire information from the choice of the peer (social learning)[7] and because one's utility from possessing an asset may depend directly on the possession of that asset by another individual (social utility).[8] Although it is virtually impossible to investigate separately those two mechanisms with non-experimental data, our paper presents novel evidence that it is not in contrast with any of them. If one considers the learning mechanism, then our paper reveals that agents learn more from peers they trust. A social utility -based interpretation instead suggests that long-lasting (and hence trustworthy) social contacts are the relevant reference group. If a conformism (herding) behavior or conspicuous consumption is driving the purchase of financial products (i.e., if it is the behavior relative to the peers that matters), then it is important to understand who the peers are with whom each individual is compared to.

Our paper is organized as follows.

We begin by describing our data in Section 3.2. Section 3.3 presents our empirical model and identification strategy, whereas Sectin 3.4 discusses our main estimation results. We collect some additional evidence in Section 3.5. In Section

[7]Theoretical models of herding and asset-price bubbles focus on learning from peers' choices (see, Bikhchandani and Sharma; 2000; Chari and Kehoe; 2004).

[8]A number of paper consider the "keeping up with the Joneses" hypothesis in explaining stock market behavior (most notably, Gali; 1994; Abel; 1990; Campbell and Cochrane; 1999).

3.6 we use simulation experiments to show the implications of social interactions for the adoption of financial products. Section 3.7 concludes.

3.2 Data description

Our analysis is made possible by the use of a unique database on friendship networks from the National Longitudinal Survey of Adolescent to Adult Health (AddHealth).[9] The AddHealth survey has been designed to study the impact of the social environment (i.e. friends, family, neighborhood and school) on students' behavior in the United States by collecting data on students in grades 7-12 from a nationally representative sample of roughly 130 private and public schools in the years 1994-1995 (Wave I). Every student attending the sampled schools on the interview day was asked to complete a questionnaire (*in-school data*) containing questions on respondents' demographic and behavioral characteristics, education, family background and friendship. A subset of students selected from the rosters of the sampled schools - about 20,000 individuals - was then asked to complete a longer questionnaire containing more sensitive individual and household information (*in-home and parental data*). Those subjects were interviewed again in 1995–1996 (Wave II), in 2001–2002 (Wave III), and in 2007-2008 (Wave IV).

From a network perspective, the most interesting aspect of the AddHealth data is the friendship information, which is based upon actual friend nominations. Indeed, students were asked to identify their best friends from a school roster (up to five males and five females).[10] This information is collected in Wave I and

[9]This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

[10]The limit in the number of nominations is not binding (even by gender). Less than 1% of

one year after, in Wave II. As a result, one can reconstruct the whole geometric structure of the friendship networks and their evolution, at least in the short run. About 10% of the nominations in our data are not reciprocal, that is there are cases where agent i nominates agent j as best friend but agent j does not list agent i among her/his best friends. We consider two agents to be connected if at least one has nominated the other as best friend. Indeed, even if agent j does not nominate i as best friend, it is reasonable to think that social interactions have taken place.[11] Such detailed information on social interaction patterns allows us to measure the peer group more precisely than in previous studies by knowing exactly who nominates whom in a network (i.e. who interacts with whom in a social group).

Moreover, one can distinguish between *strong* and *weak* ties in the data. We define a *strong* tie or relationship between two students if they have nominated each other in both waves (i.e. in Wave I in 1994-1995 *and* in Wave II in 1995-1996) and a *weak* tie or relationship if they have nominated each other in one wave only (Wave I *or* Wave II).

The information about financial decisions is collected in Wave III. Unfortunately, friends' nominations are not collected in this wave, as some individuals have left high school. However, more than 80% are still at school and the large majority of the individuals (more than 75%) declare that they are still in contact with at least one friend nominated in the past wave. Of course, new friends can be created at the time of Wave III, and/or friendship relationships between schoolmates may change over time (see Section 3.3.4). The network of social contacts during high school remains however a good approximation of face-to-face information they are (or have been in a recent past) exposed to. The questionnaire of Wave III contains detailed information on the use of financial and payment

the students in our sample show a list of ten best friends, both in Wave I and Wave II.

[11] An alternative definition of network link that exploits the direction of the nominations does not substantially change our results.

instruments like saving and checking accounts, credit cards, loans, shares of stock in publicly held corporations, mutual funds, or investment trusts. Table 3.4 reports on the financial activity participation of the agents in our sample. More than 60% of the students have a checking account, a saving account, and a credit card. About 40% have a credit card debt and more than 30% has a student loan. Interestingly, 25% of individuals own shares of stock in publicly held corporations, mutual funds, or investment trusts, including stocks in IRAs. For each individual, we construct an index of financial activity participation using a traditional principal component analysis, where the loadings of these different activities are used to derive a total score. Our measure of financial activity is the first principal component. It explains over one-third of the total inertia.[12] The last column of Table 3.4 shows that each financial activity is positively correlated with this variable, meaning that the larger the variety of financial products that an individual uses, the higher the value of our indicator of financial participation. The index ranges between 0 and 2.64, with mean equal to 1.47 and substantial variation around this mean value (standard deviation equal to 0.77).

A unique feature of our data is that, by matching the identification numbers of the friendship nominations to respondents' identification numbers, one can obtain information on all nominated friends. Such a data structure thus allow us to investigate the role of peers' adoption of financial instruments on individual decisions.[13]

Before proceeding with the formal analysis, we provide a heuristic description of a social network to illustrate the relationship between financial activity and the network topology. Figure 3.1 shows a representative network. Each node

[12]PCA uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables (called principal components). This transformation is defined in such a way that the first principal component accounts for the largest portion of variability in the data.

[13]The other existing surveys that report friends' nomination are ego-networks, i.e. the respondent lists her contacts and some basic characteristics of them (such as gender, education, employment status). Detailed information about nominated contacts is typically not available.

represents an agent, with the size of the node proportional to her/his level of participation to the financial market. The lines represent the connections between the agents; the thicker they are, the longer the interaction relationship between pairs of agents. As can be seen from the picture, agents in more cohesive groups characterized by a relatively high density of ties tend to show a higher and more similar level of financial activity. This stylized fact motivates our analysis in the following sections.

The sample of individuals that are followed over time and have non-missing information for our target variables both in Waves I, II and III consists of 12,874 individuals. As is common with AddHealth data, a further reduction in sample size is due to the network construction procedure - roughly 20% of the students do not nominate any friends and another 20% cannot be correctly linked.[14] In addition, in this study we focus on networks with size between 10 to 50 agents to cope with the computational burden required by the use of Bayesian estimation procedures. Our final sample consists of 569 individuals distributed over 21 networks.[15] Nevertheless, we report in Appendix E the results which are obtained using traditional estimation techniques on the more extensive sample. They remain qualitatively unchanged.[16] Table A1 in Appendix A gives precise definitions of the variables used in our study as well as their descriptive statistics.[17] Table A1 in Appendix A also gives details on nomination data. The mean and the standard deviation of network size are roughly 27 and 13 students, respectively. On average, these individuals have 23% strong ties and 76% weak ties.[18]

[14]The representativeness of the sample is preserved. Summary statistics are available upon request.

[15]Our results, however, do not depend crucially on these network size thresholds. They remain qualitatively unchanged when changing the network size window slightly.

[16]Even in this case we do not consider networks at the extremes of the network size distribution (i.e. consisting of 2-3 individuals or more than 400) because peer effects can show extreme values (too high or too low) in these edge networks (see Calvo-Armengol et al., 2009).

[17]By comparing those summary statistics with the ones of the original sample, it appears that the representativeness of the sample is preserved. Results available upon request.

[18]Information at the school level, such as school quality and the teacher/pupil ratio, is also available. We do not use it since our sample of networks is within schools and we include fixed network effects in our estimation strategy.

3.3 Empirical model and estimation strategy

3.3.1 The network model

Consider a population of n individuals partitioned into $\bar{r}$ networks. For the n_r individuals in the rth network, their connections with each other are represented by an $n_r \times n_r$ adjacency matrix $\mathbf{G}_r = [g_{ij,r}]$ where $g_{ij,r} = 1$ if individuals i and j are friends and $g_{ij,r} = 0$ otherwise.[19] Let $\mathbf{G}_r^* = [g_{ij,r}^*]$ be the row-normalized $\mathbf{G}_r$ such that $g_{ij,r}^* = g_{ij,r} / \sum_{k=1}^{n_r} g_{ik,r}$.

The financial activity of individual i in network r, $y_{i,r}$, is given by

$$y_{i,r} = \phi \sum_{j=1}^{n_r} g_{ij,r} y_{j,r} + \sum_{k=1}^{p} x_{ik,r}\beta_k + \sum_{k=1}^{p} (\sum_{j=1}^{n_r} g_{ij,r}^* x_{jk,r})\delta_k + \eta_r + \epsilon_{i,r}. \quad (3.1)$$

In this model, $\sum_{j=1}^{n_r} g_{ij,r} y_{j,r}$ denotes the aggregate financial activity of i's direct contacts with its coefficient ϕ representing *the endogenous effect*, wherein an individual's choice may depend on those of his/her contacts about the same activity.[20] $x_{ik,r}$ indicates the k_{th} exogenous variable accounting for observable differences in individual characteristics (e.g. gender, race, education, income, family background, etc.). $\sum_{j=1}^{n_r} g_{ij,r}^* x_{jk,r}$ is the average value of the exogenous variables over i's direct contacts with its coefficient δ_k representing *the contextual effect*, wherein an individual's financial activity index may depend on the exogenous characteristics of his/her contacts. η_r is a network-specific parameter representing *the correlated effect*, wherein individuals in the same group tend to behave similarly because they face a common environment. $\epsilon_{i,r}$ is an i.i.d. error term with zero mean and finite variance σ^2.

Model (4.1) can be extended to the case of heterogeneous peer effects. If we

[19] For ease of presentation, we focus on the case where the connections are undirected and no agent is isolated so that G_r is symmetric and $\sum_{j=1}^{n} g_{ij,r} \neq 0$ for all i.

[20] Given we are modeling the financial activity of agents it seems more appropriate to consider an "aggregate" model instead of an "average" one. The first type of models allows the number of peers to be relevant in shaping the agents' activity, while the second do not consider this information, i.e. it uses average values of peers' activity (see, Liu et al.; 2011).

consider that each "ego-network" (i.e. the social contacts of a specific agent) can be split into two different types (weak and strong ties), then Model (4.1) becomes

$$\begin{aligned} y_{i,r} &= \phi^S \sum_{j=1}^{n_r} g^S_{ij,r} y_{j,r} + \phi^W \sum_{j=1}^{n_r} g^W_{ij,r} y_{j,r} + \boldsymbol{x}'_{i,r}\beta \\ &+ \frac{1}{g^S_{i,r}} \sum_{j=1}^{n_r} g^S_{ij,r} x'_{j,r}\delta^S + \frac{1}{g^W_{i,r}} \sum_{j=1}^{n_r} g^W_{ij,r} x'_{j,r}\delta^W + \eta_r + \epsilon_{i,r}, \end{aligned} \tag{3.2}$$

where $g^S_{i,r} = \sum_{j=1}^{n} g^S_{ij,r}$ and $g^W_{i,r} = \sum_{j=1}^{n} g^W_{ij,r}$ are the total number of strong and weak ties each individual i has in network r. In this model, ϕ^S and ϕ^W represent *the endogenous effects* (i.e. the effect of strong and weak ties' financial activity on one's own financial choices respectively) while δ^S and δ^W capture the impact of the exogenous characteristics of the peers - which are allowed to have a varying effect by peer-type.

3.3.2 Identification and estimation

A number of papers have dealt with the identification and estimation of peer effects with network data (see, e.g. Bramoullé et al.; 2009; Liu and Lee; 2010; Calvó-Armengol et al.; 2009; Lin; 2010b; Lee et al.; 2010). Below, we review the crucial issues while explaining how we tackle them.

Reflection problem In linear-in-means models, simultaneity in the behavior of interacting agents introduces a perfect collinearity between the expected mean outcome of the group and its mean characteristics. Therefore, it is difficult to differentiate between the effect of peers' choice of effort (*endogenous effects*) and peers' characteristics (*contextual effects*) that do have an impact on their effort choice (the so-called reflection problem; Manski; 1993). Basically, the reflection problem arises because, in the standard approach, individuals interact in groups - individuals are affected by all individuals belonging to their group and by nobody outside the group. In the case of social networks, instead, this is nearly never true

since the reference group is individual specific. For example, take individuals i and k such that $g_{ik,r} = 1$. Then, individual i is directly influenced by $\overline{y}_i = \sum_{j=1}^{n_r} g_{ij,r} y_j$ while individual k is directly influenced by $\overline{y}_k = \sum_{j=1}^{n_r} g_{kj} y_j$, and there is little chance for these two values to be the same unless the network is complete (i.e. everybody is linked with everybody).[21]

Correlated effects While a network approach allows us to distinguish endogenous effects from correlated effects, it does not necessarily estimate the causal effect of peers' influence on individual behavior. The estimation results might be flawed because of the presence of peer-group specific *unobservable* factors affecting both individual and peer behavior. For example, a correlation between the individual and the peer-school performance may be due to an exposure to common factors (e.g. having good teachers) rather than to social interactions. The way in which this has been addressed in the literature is to exploit the architecture of network contacts to construct valid IVs for the endogenous effect. Since peer groups are individual specific in social networks, the characteristics of indirect friends are natural candidates. Consider the network in Figure 3.2. Individual k affects the behavior of individual i only through her/his common friend j, and she/he is not exposed to the factors affecting the peer group consisting of individual i and individual j. As a result, the characteristics x_k of individual k are valid instruments for y_j, the endogenous outcome of j.

Sorting In most cases, individuals sort into groups non-randomly. For example, students whose parents are low-educated (or worse than average in unmeasured ways) would be more likely to sort with low human capital peers. If the variables that drive this process of selection are not fully observable, potential correlations between (unobserved) group-specific factors and the target regressors are major sources of bias. The richness of social network data (where we observe individuals over networks) provides a possible way out by the use of *network fixed*

[21] Formally, social effects are identified (i.e. no reflection problem) if I, G, G^2 and G^3 are linearly independent, where $G = diag(G_r)_{r=1,..\bar{r}}$.

effects. Network fixed effects are a remedy for the selection bias that originates from the possible sorting of individuals with similar unobserved characteristics into a network. The underlying assumption is that such unobserved characteristics are common to the individuals within each network. This is reasonable in our case study where the networks are quite small (see Section 3).

However, if there are individual-level unobservables that drive both network formation and outcome choice, this strategy fails. For example, one can envision the existence of unobservable (or unmeasurable) factors, such as risk aversion or optimism, which are possibly relevant both in social contexts and for financial decisions making. Recently, Goldsmith-Pinkham and Imbens (2013) and Hsieh and Lee (2011) highlight the fact that endogeneity of this sort can be included in the model. Individual-level correlated unobservables would motivate the use of parametric modeling assumptions and Bayesian inferential methods to integrate a network formation with the study of behavior over the formed networks. The next section contains the results which are obtained by applying this approach to our case.

3.3.3 Endogenous Network Formation

Goldsmith-Pinkham and Imbens (2013) and Hsieh and Lee (2011) propose two slightly different ways to estimate peer effects with unobservables driving both link formation and outcome.[22] In Goldsmith-Pinkham and Imbens (2013) unobservables are dichotomous and only one network is considered. As we have multiple networks in our data, we follow Hsieh and Lee (2011).[23] They present a model with one peer-type - which correspond to Model (4.1). We implement

[22]The Bayesian approach allows to model couple-specific unobserved heterogeneity, for each possible couple in the sample. A traditional Heckman-type selection model is configured to capture individual-specific unobserved heterogeneity. The inclusion of alter heterogeneity would imply the computation of high-dimensional multivariate normal integrals, which is unfeasible using standard methods.

[23]Another difference between those two procedures is that Goldsmith-Pinkham and Imbens (2013) set the same unobservable in both link formation and outcome equation while Hsieh and Lee (2011) use different unobservables for those equations and let them to be correlated.

here an extension of their method to the case of heterogeneous peer effects. If there is an unobservable characteristic that drives the choice of, say, strong ties and is correlated with $\epsilon_{i,r}$ then $g^S_{ij,r}$ is endogenous - estimates of Model (3.2) are biased. By failing to account for similarities in (unobserved) characteristics, similar behaviors might mistakenly be attributed to peer influence when they simply result from similar characteristics. Let $z_{i,r}$ denote such an *unobserved characteristic* which influence the link formation process. Let us also assume that $z_{i,r}$ is correlated with $\epsilon_{i,r}$ in Model (3.2) according to a bivariate normal distribution

$$(z_{i,r}, \epsilon_{i,r}) \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_z^2 & \sigma_{\varepsilon z} \\ \sigma_{\varepsilon z} & \sigma_\varepsilon^2 \end{pmatrix}\right). \tag{3.3}$$

Agents choose social contacts at two points in time, t-1 and t. At each time, agent i chooses to be friends with j according to a vector of observed and unobserved characteristics in a standard link formation probabilistic model

$$P(g_{ij,r,t-1} = 1 | x_{ij,r}, z_{i,r}, z_{j,r}, \gamma_{t-1}, \theta_{t-1}) = \Lambda(\gamma_{0,t-1} + \sum_k |x_{i,r} - x_{j,r}|\gamma_{k,t-1} + |z_{i,r} - z_{j,r}|\theta_{t-1}), \tag{3.4}$$

and

$$P(g_{ij,r,t} = 1 | x_{ij,r}, z_{i,r}, z_{j,r}, g_{ij,r,t-1}, \gamma_t, \theta_t, \lambda) = \Lambda(\gamma_{0,t} + \lambda g_{ij,r,t-1} + \sum_k |x_{i,r} - x_{j,r}|\gamma_{k,t} + |z_{i,r} - z_{j,r}|\theta_t), \tag{3.5}$$

where $\Lambda(\cdot)$ is a logistic function. Homophily behavior in the unobserved characteristics implies that $\theta_\tau < 0$, where $\tau = t - 1, t$, this meaning that the closer two individuals are in terms of unobservable characteristics, the higher is the probability that they are friends. The same argument holds for observables. If $\sigma_{\varepsilon z}$ and θ_τ are different from zero, then networks $g^S_{ij,r}$ and $g^W_{ij,r}$ in model (4.1) are

endogenous. From Model (3.4) - (3.5), the probability of observing a weak tie is

$$P(g^W_{ij,r} = 1|x_{ij,r}, z_{i,r}, z_{j,r}, \gamma_t, \theta_t, \lambda, \gamma_{t-1}, \theta_{t-1})$$

$$= P(g_{ij,r,t} = 1|x_{ij,r}, z_{i,r}, z_{j,r}, \gamma_t, \theta_t, \lambda, g_{ij,r,t-1} = 0) \times P(g_{ij,r,t-1} = 0|x_{ij,r}, z_{i,r}, z_{j,r}, \gamma_{t-1}, \theta_{t-1})$$

$$+P(g_{ij,r,t} = 0|x_{ij,r}, z_{i,r}, z_{j,r}, \gamma_t, \theta_t, \lambda, g_{ij,r,t-1} = 1) \times P(g_{ij,r,t-1} = 0|x_{ij,r}, z_{i,r}, z_{j,r}, \gamma_{t-1}, \theta_{t-1})$$

whereas the probability of observing a strong tie is

$$P(g^S_{ij,r} = 1|x_{ij,r}, z_{i,r}, z_{j,r}, \gamma_t, \theta_t, \lambda, \gamma_{t-1}, \theta_{t-1})$$

$$= P(g_{ij,r,t} = 1|x_{ij,r}, z_{i,r}, z_{j,r}, \gamma_t, \theta_t, \lambda, g_{ij,r,t-1} = 1) \times P(g_{ij,r,t-1} = 1|x_{ij,r}, z_{i,r}, z_{j,r}, \gamma_{t-1}, \theta_{t-1}).$$

In this way, we have modeled the probability of being a strong or weak ties including unobservables that are allowed to be correlated with the error term in the outcome equation.[24] Joint normality in (3.3) implies $E(\epsilon_{i,r}|z_{i,r}) = \frac{\sigma_{\varepsilon z}}{\sigma_z^2} z_{i,r}$, when conditioning on $z_{i,r}$. Hence, the outcome equation is

$$\begin{aligned} y_{i,r} &= \phi^S \sum_{j=1}^{n_r} g^S_{ij,r} y_{j,r} + \phi^W \sum_{j=1}^{n_r} g^W_{ij,r} y_{j,r} + \boldsymbol{x}'_{i,r}\beta + \frac{1}{g^S_{i,r}} \sum_{j=1}^{n_r} g^S_{ij,r} x'_{j,r}\delta^S \\ &+ \frac{1}{g^W_{i,r}} \sum_{j=1}^{n_r} g^W_{ij,r,t} x'_{j,r}\delta^W + \eta_r + \frac{\sigma_{\varepsilon z}}{\sigma_z^2} z_{i,r} + u_{i,r}, \end{aligned} \tag{3.6}$$

where $u_{i,r} \sim N(0, \sigma_z^2 - \frac{\sigma_{\varepsilon z}^2}{\sigma_z^2})$. Note that if no correlation is at work ($\sigma_{\varepsilon z} = 0$), then estimating equation (3.6) or (3.2) is equivalent. Given the complexity of this framework, it is convenient to simultaneously estimate the parameters of equations (3.4), (3.5) and (3.6) with a Bayesian approach. Bayesian inference requires the computation of marginal distribution for all parameters. However, since this requires integration of complicated distributions over several variables, a closed form solution is not readily available and Markov Chain Monte Carlo

[24] The procedure can be easily extended to include more than one unobservable factor.

(MCMC) techniques are usually employed to obtain random draws from posterior distributions. The unobservable variable ($z_{i,r}$) is thus generated according to the joint likelihood of link formation and outcome - it is drawn in each MCMC step together with the parameters of the model. The Gibbs sampling algorithm allows us to draw random values for each parameter from their posterior marginal distribution, given previous values of other parameters. Once stationarity of the Markov Chain has been achieved, the random draws can be used to study the empirical distributions of the posterior.[25]

3.3.4 Network topology misspecification

Traditional estimation methods of network models hinges upon the correct specification of the topology of network contacts (as captured by the adjacency matrix G).

There are two possible sources of misspecification: *(i)* missing nodes and *(ii)* mispecified links.

The first relates to the definition of each agent's set of peers. In our application, we observe (and make inference on) the connection profile of each agent at time t=1,2, while the financial decisions are made at time t=3. One can thus argue that at time t=3 there are agents that are not in our sample but influence an individual's financial decisions. An example would be new social ties giving financial advice at time t=3. The natural question is thus: how would the presence of unobservable agents affect our estimation results?

The second stems from the fact that the connections among observed agents may be misspecified. In our application, some individuals may not report at time t=1,2 some friends who are relevant for her/his future financial decisions or an observed link at time t=1,2 may not exit anymore at the time the financial decision

[25]See Appendix B for more details on the estimation procedure. An introduction to Monte Carlo methods in Bayesian econometrics can be found in Chib (1996) and Robert and Casella (2004).

is made (t=3). Therefore, the question here is: how does link misspecification affect our estimation results?

In this section, we shall show that the bayesian approach that we propose in Section 3.3.3 is able to take into account a possible network topology misspecification.

(i) Missing nodes

Suppose we have two types of agents in our network, observables (O) and unobservables (U). Model (4.1) can be written as

$$y_{i,r} = \phi_o \sum\nolimits_{j=1,j\in O}^{n_r} g_{ij,r}y_{j,r} + \phi_{ou} \sum\nolimits_{j=1,j\in U}^{n_r} g_{ij,r}y_{j,r} + \sum\nolimits_{k=1}^{p} x_{ik,r}\beta_k + \epsilon_{i,r}, \forall i \in O;$$

$$y_{i,r} = \phi_{uo} \sum\nolimits_{j=1,j\in O}^{n_r} g_{ij,r}y_{j,r} + \phi_u \sum\nolimits_{j=1,j\in U}^{n_r} g_{ij,r}y_{j,r} + \sum\nolimits_{k=1}^{p} x_{ik,r}\beta_k + \epsilon_{i,r}, \forall i \in U.$$

where ϕ_o is the intra-observed agents peer effect, ϕ_u is the intra-unobserved agents peer effect, ϕ_{ou} and ϕ_{uo} are cross-groups peer effcts - w.l.o.g. we omitt the contextual effects and the network fixed effects in order to avoid cumbersome notation. In matrix form, we have that

$$\mathbf{Y}_O = \phi_o \mathbf{G}^O \mathbf{Y}_O + \phi_{ou} \mathbf{G}^{OU} \mathbf{Y}_U + \mathbf{X}_O \beta + \epsilon_O,$$

$$\mathbf{Y}_U = \phi_{uo} \mathbf{G}^{UO} \mathbf{Y}_O + \phi_u \mathbf{G}^U \mathbf{Y}_U + \mathbf{X}_U \beta + \epsilon_U.$$

where G^O has $\{g_{ij}, i \in O, j \in O\}$ as a generic element, G^U has $\{g_{ij}, i \in U, j \in U\}$ as a generic element, G^{OU} has $\{g_{ij}, i \in O, j \in U\}$ as a generic element and G^{UO} has $\{g_{ij}, i \in U, j \in O\}$ as a generic element. The empirical model that we estimate

is

$$\mathbf{Y}_O = \phi_o \mathbf{G}^O \mathbf{Y}_O + \mathbf{X}_O \beta + \epsilon_O^*,$$

where

$$\begin{aligned} \epsilon_O^* &= \epsilon_O + \phi_{ou} \mathbf{G}^{OU} \mathbf{Y}_U = \\ &= \epsilon_O + \phi_{ou} \mathbf{G}^{OU} \mathbf{M}_U (\phi_u \mathbf{G}^{UO} \mathbf{J}^O \mathbf{X}_O \beta + \mathbf{X}_U \beta) \end{aligned} \tag{3.7}$$

with $\mathbf{M}_U = (I - \phi_u \mathbf{G}^U - \phi_{ou}\phi_{uo}\mathbf{C}^U)^{-1}$, $\mathbf{J}^O = (I - \phi_o \mathbf{G}^O)^{-1}$ and $\mathbf{C}^U = \mathbf{G}^{UO}\mathbf{G}^O\mathbf{G}^{OU}$. As a result, it may be that $E(\mathbf{G}^O \mathbf{Y}_O{}' \epsilon_O^*) \neq 0$.

The instrumental variables traditionally adopted in the estimation of network models, i.e. the characteristics of friends of friends, may not be valid.

Specifically,

$$\widehat{\theta}_{2sls} = \left(\mathbf{X}_O' \mathbf{P}_\mathbf{K} \mathbf{X}_O\right)^{-1} \mathbf{X}_O' \mathbf{P}_\mathbf{K} \mathbf{Y}_\mathbf{O},$$

where $\mathbf{P}_\mathbf{K} = \mathbf{Q}_\mathbf{K}(\mathbf{Q}_\mathbf{K}{}'\mathbf{Q}_\mathbf{K})^{-1}\mathbf{Q}_\mathbf{K}{}'$, and $\mathbf{Q}_\mathbf{K} = [\mathbf{G^O}, (\mathbf{G^O})^2, (\mathbf{G^O})^3, ..., (\mathbf{G^O})^K]\mathbf{X}_O$ are the optimal IV for the endogenous variable. If the error term is (3.7), then it may be that $E(\mathbf{Q}_\mathbf{K}{}' \epsilon_O^*) \neq 0$. The intuition is that the error term (3.7) contains the characteristics of the observed nodes that are used as instruments.[26]

The bias takes the following form

$$B_{2sls} = E[\left(\mathbf{X}_O' \mathbf{P}_\mathbf{K} \mathbf{X}_O\right)^{-1} \mathbf{X}_O' \mathbf{P}_\mathbf{K} \phi_{ou} \mathbf{G}^{OU} \mathbf{Y}_\mathbf{U}].$$

It will be different from zero if *(i)* $\phi_{ou} \neq 0$ and/or *(ii)* $\phi_{uo} \neq 0$, which mean no interactions between observable and unobservable friends and/or *(iii)* $\text{tr}\left(\mathbf{G}^{OU}\mathbf{M}_U\mathbf{G}^{UO}\mathbf{J}^O\right) \neq 0$, which means that there is at least one unobservable

[26]To be more precise, in the error term the characteristics of observables nodes X_O are multiplied by $\phi_{ou} G^{OU} M_U \phi_u G^{UO} J^O \beta$ i. Therefore the correlation of the error term with the instrument depends on the structure of the entire network and on the interaction parameters.

friend that connects two observable friends.

Our Bayesian approach would not be affected by this bias. It is not based on IVs and it considers the presence of unobervable factors correlated with the endogenous variable. Such unobservable factors may include those arising from the presence of missing nodes.

In order to see more clearly why this is the case, let us assume that z as defined in (3.3) is

$$z = s + \phi_{ou}\mathbf{G}^{OU}\mathbf{Y}_U$$

where s is a vector capturing other unobservables characteristics (which can also be zero). Let us denote $z = s + \widetilde{x} + \widetilde{\varepsilon}$, where $\widetilde{x} = \phi_{ou}\mathbf{G}^{OU}\mathbf{M}_U(\phi_u\mathbf{G}^{UO}\mathbf{J}^O\mathbf{X}_O\beta + \mathbf{X}_U\beta)$ and $\widetilde{\varepsilon} = \phi_{ou}\mathbf{G}^{OU}\mathbf{M}_U(\phi_u\mathbf{G}^{UO}\mathbf{J}^O\epsilon_O + \epsilon^U)$. For the sake of clarity and w.l.o.g., suppose that $E(s) = -\widetilde{x}$. The distributional assumption (3.3) implies that

$$\sigma_{\varepsilon z} = \sigma_{\varepsilon s} + \ \phi_{ou}\phi_u \ \sigma_\varepsilon \text{tr}\left(\mathbf{G}^{OU}\mathbf{M}_U\mathbf{G}^{UO}\mathbf{J}^O\right)$$

and $\sigma_z^2 = \ \sigma_s^2 + \sigma_{\widetilde{\varepsilon}}^2 - 2\sigma_{\widetilde{\varepsilon}s}$, where $\sigma_{\widetilde{\varepsilon}} = \sigma_\varepsilon \text{tr}\,(\Omega)$, $\Omega = \phi_{ou}^2\phi_{uo}^2 \text{tr}\left((\mathbf{G}^{OU}\mathbf{M}_U\mathbf{G}^{UO}\mathbf{J}^O)'(\mathbf{G}^{OU}\mathbf{M}_U\mathbf{G}^{UO}\mathbf{J}^O)\right)$.

$\sigma_{\varepsilon z}$ captures the correlation due to the presence of missing nodes. It would be zero is any of the conditions *(i)*, *(ii)* and *(iii)* would be zero.

A comparison between the traditional IV estimator and the bayesian estimator should then be informative on the importance of missing node effects.

In our application, we find no significant difference (see Table 3.5, columns 6 and 7). Going back to condition *(i)*, *(ii)* and *(iii)* it is very likely for example that even if there are relevant missing friends (i.e. ϕ_{ou} and $\phi_{ou} \neq 0$), those missing friends are not common to any observed individual, given that after they leave school at wave III they are very likely in different environments.

(ii) Mispecified links

Suppose we observe all the agents but connections among them can be misspecified. This is likely to happen in our application since friendship relationships are collected in a time different from when financial choices are taken. Assume that the real DGP is

$$\mathbf{Y} = \phi\mathbf{\Gamma Y} + \mathbf{X}\beta + \epsilon,$$

where $\mathbf{\Gamma}$ is the real adjacency matrix, with generic element equal to γ_{ij}, which is unknown to the econometrician. We observe

$$\mathbf{Y} = \phi\mathbf{GY} + \mathbf{X}\beta + \mathbf{e},$$

where $\mathbf{G}$ is the observed adjacency matrix, with generic element equal to g_{ij}. We can write $\mathbf{G} = \mathbf{\Gamma} - \mathbf{\Theta} + \mathbf{\Delta}$, where $\mathbf{\Theta}$ is a matrix, with generic element equal to θ_{ij} taking value one if the ij_{th} link is not observed but exists (i.e. $g_{ij} = 0$ and $\gamma_{ij} = 1$) and zero otherwise; $\mathbf{\Delta}$ is a matrix with generic element equal to δ_{ij} taking value one if the ij_{th} link is observed but does not exist (i.e. $g_{ij} = 1$ and $\gamma_{ij} = 0$) and zero otherwise. Equation (3.3.4) becomes

$$\begin{aligned} \mathbf{Y} &= \phi(\mathbf{G} + \mathbf{\Theta} - \mathbf{\Delta})\mathbf{Y} + \mathbf{X}\beta + \epsilon \\ &= \phi\mathbf{GY} + \mathbf{X}\beta + \xi, \end{aligned} \tag{3.8}$$

where

$$\begin{aligned} \xi &= \epsilon + \phi(\mathbf{\Theta} - \mathbf{\Delta})\mathbf{Y} \\ &= \epsilon + \phi(\mathbf{\Theta} - \mathbf{\Delta})\mathbf{M}\mathbf{X}\beta + \phi(\mathbf{\Theta} - \mathbf{\Delta})\mathbf{M}\epsilon \end{aligned} \tag{3.9}$$

with $\mathbf{M} = (I - \phi\mathbf{\Gamma})^{-1}$. Thus, as in the previous case, it may be that $E(\mathbf{G}\mathbf{Y}'\xi) \neq 0$.[27]

Also in this case the instrumental variables traditionally adopted in the estimation of network models may not be valid. We have

$$\widehat{\theta}_{2sls} = \left(\mathbf{X}'\mathbf{P}_{\mathbf{K}}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{P}_{\mathbf{K}}\mathbf{Y},$$

where $\mathbf{P}_{\mathbf{K}} = \mathbf{R}_{\mathbf{K}}(\mathbf{R}_{\mathbf{K}}'\mathbf{R}_{\mathbf{K}})^{-1}\mathbf{R}_{\mathbf{K}}'$, and $\mathbf{R}_{\mathbf{K}} = [\mathbf{G}, \mathbf{G}^2, \mathbf{G}^3, ..., \mathbf{G}^K]\mathbf{X}$ are the optimal IV for the endogenous variable. If the error term is (3.10), then it may be that $E(\mathbf{R}_{\mathbf{K}}'\xi) \neq 0$.

Our Bayesian approach would instead provide a consistent estimator.[28] The unobservable factor z in (3.3) can be written as

$$z = s + \phi(\mathbf{\Theta} - \mathbf{\Delta})\mathbf{Y}$$

where s is a vector capturing other unobservables characteristics (which can also be zero). Let us denote $z = s + \widetilde{x} + \widetilde{\varepsilon}$, where $\widetilde{x} = \phi(\mathbf{\Theta} - \mathbf{\Delta})\mathbf{M}\mathbf{X}\beta$ and $\widetilde{\varepsilon} = \phi(\mathbf{\Theta} - \mathbf{\Delta})\mathbf{M}\epsilon$.

Assuming w.l.o.g. that $E(s) = -\widetilde{x}$, we have

$$\sigma_{\varepsilon z} = \sigma_{\varepsilon s} + \ \phi\ \sigma_{\varepsilon}\mathrm{tr}\left((\mathbf{\Theta} - \mathbf{\Delta})\mathbf{M}\right).$$

$\sigma_{\varepsilon z}$ captures the correlation due to the presence of missing links. It would be zero if *(i)* $\mathbf{\Theta} = \mathbf{0}$ and $\mathbf{\Delta} = \mathbf{0}$, which means no mispecification of the adjacency matrix and/or *(ii)* $\mathrm{tr}\left((\mathbf{\Theta} - \mathbf{\Delta})\mathbf{M}\right) = 0$, which means that there are no loops that include mispecifiefd links. Therefore, also in this case a comparison between the

[27] As mentioned in footnote 26, the correlation of the error term with the instrument depends on the structure of the entire network and on the interaction parameters.

[28] In a Bayesian estimation, consistency means that the posterior probability of the parameter is concentrated around the true value as the sample size increases, assuming that the true value belongs to the parameter space being considered.

traditional IV estimator and the bayesian estimator should be informative on the importance of link mispecification.

3.4 Estimation results

The aim of our empirical analysis is twofold: (*i*) to assess the presence of peer effects in the adoption of financial products, (*ii*) to differentiate between the impact of weak and strong social ties.

3.4.1 Peer effects

Table 3.5 collects the estimation results of model (4.1), that is without distinguishing between strong and weak ties. Columns (1) to (6) report the results when network exogeneity is assumed, with different estimation methods. Column (7) shows the Bayesian estimation results, which account for a possible network endogenity. Columns (1) to (3) report OLS estimates with an increasing set of controls. Column (1) includes individual socio-demographic characteristics (age, race, gender, education, employment status, occupation, parental education, marital status, family background variables, etc.), while column (2) extends the number of control variables to include peers' characteristics. This specification addresses the concern that a correlation between own and peers' behavior is simply driven by similar (observable) characteristics between peers. Finally, column (3) adds network fixed effects, thus accounting for any further unobserved factors common to all individuals in a social group. The issue addressed here is that correlated actions between connected agents may be simply driven by common shocks or sorting into groups according to network-specific unobserved characteristics. Column (4) presents the estimation results using ML, that is when the

simultaneity which is endemic in spatial models is accounted for.[29] Columns (5) and (6) are devoted to the IV estimates. As explained in Section 3.3.2, the IV strategy that is now standard in network model estimation consists of exploiting network architecture and uses peers of peers' characteristics as instruments for peers' behavior. Table 3.6 reports the first stage results. The F-statistic confirms the relevance of the IVs. Because of the many-IVs bias that may arise in estimating spatial models with IVs, we follow Liu and Lee (2010) and also use a bias-corrected IV.[30] Finally, column (7) reports means and standard deviations of the posterior distributions of the parameters of Model (3.4) - (3.5) - (3.6), that is with correlated unobservables, estimated by Bayesian methods. We let our Markov Chain run for 80,000 iterations, discarding the first 7,000, even though ergodicity of the Markov Chain is achieved quite quickly. It appears that the Bayesian estimates (column (7)) are remarkably similar to the ones that are obtained using the IV biased-corrected (column (6)). This suggests that unobservable factors influencing the link formation are not relevant in the financial decisions of agents. Indeed, the estimated correlation between unobservables in the outcome and link formation equations ($\sigma_{\varepsilon z}$) is not significantly different from zero. For completeness, Figures 3.3 and 3.4 show the kernel density estimates of the posterior distributions (left panel) and the Markov chain (right panel) of ϕ and $\sigma_{\varepsilon z}$. The time-series of the values of the chains (right panel) reveals that stationarity has been achieved.[31] Table 3.5 shows that the coefficients are quite stable across columns.[32] It reveals that the effect of peers' financial activity on

[29]Spatial models are simultaneous equation models where peers' behavior depends on own behavior. This implies that $\sum_{j=1}^{n_r} g_{ij,r} y_{j,r}$ is correlated with the error term $\varepsilon_{i,r}$ in equation (4.1). ML accounts for this simultaneity as it is based on the reduced form. Network fixed effects cannot be included in the model because the group mean $\overline{y}_r$ is not a sufficient statistics for η_r when the adjacency matrix is not row-normalized, see Lee et al. (2010).

[30]See Appendix C for more details. For the sake of brevity, the appendix focuses on the case with weak and strong ties. The case with one peer effect is just a special case, that is when $\phi^S = \phi^W$.

[31]The kernel densities and the time-series of the values of the chain for the parameters of the control variables are reported in Appendix D, Figure D1-D3.

[32]The estimate of peer effects (ϕ) using OLS is not surprisingly upper-biased. The IV estimates also suffer from a bias due to the large number of IVs, that are employed when estimating

own activity is significant and positive, i.e. there are *peer effects in financial activity*. When observable and unobservable characteristics are controlled for (estimates in column 7), in an average group of four agents, a standard deviation increase in the level of financial activity of each of the peers translates into a roughly 22% increase of a standard deviation in the individual's financial activity. In terms of the different financial activities embedded in the composite index, the estimate implies an increase of about 9% in the individual probability of getting a credit card, 6% in the probability of opening a checking or saving account, 4% in the probability of buying shares, 3% in the probability of getting a loan, and 8% in the probability of having a credit card debt.[33] These are non-negligible effects, especially given our long list of individual and peers' controls.[34] Observe that the policy maker can rarely manipulate peer outcomes. Peer effects can be seen as a mechanism through which an exogenous shock could propagate through the networks. We devote Section 3.6 to analyze these diffusion mechanisms via Monte Carlo simulations. Interestingly, parents' role does not seem to be crucial for financial activity in our sample of young people. It should be noted that the majority of individuals in our sample do not live with their parents anymore, and less than 30% are not employed.

3.4.2 Peer effects by peer-type

Table 3.7 collects the estimation results of Model (3.2). It has a structure similar to Table 3.5.[35] Column (4) shows the Bayesian estimation results, which account

a spatial model (see Appendix C)

[33]We compute these estimated probabilities using the marginal effect of an increase of the financial activity index on the probability of adopting each of the different financial products. Marginal effects are evaluated at the sample mean: $m(\beta) = \phi(\overline{x}\beta)\beta$, where $\phi(\cdot)$ is the normal probability density function. Results do not change significantly if the average of individual marginal effects is instead considered.

[34]Although the computational burden requested by the Bayesian procedure prevents us from performing this type of estimation on the entire sample, we report in Appendix E, Table E1, the OLS, ML and IV results for the entire sample.

[35]For brevity, we do not report here the ML estimation results. They are similar to the IV-bias corrected estimation results.

for possible endogeneity of strong and weak tie networks. The results in Table 3.7 do not change qualitatively across columns and reveal that the financial choices of weak ties have no significant impact on individual financial activity, while the financial choices of strong ties do have a positive and significant effect on own ones.[36] OLS and IV estimates seem to overestimate the effects. The IV bias-corrected and Bayesian estimates are very close to each other. This also means that also unobservable factors influencing the *strength* of a tie are not relevant in the financial decisions of agents ($\sigma_{\varepsilon z}$ is not significantly different from zero).[37] Given that our networks are quite small in size, it is thus likely that any correlated unobserved factor is already captured by the network fixed effects. The upper panel of Figures 3.5 shows the kernel density estimates of the posterior distributions of ϕ^S and ϕ^W. Two features of note are: (i) the distribution of ϕ^W is centered on zero; (ii) the distribution of ϕ^S is shifted towards the right.[38] This confirms that the effect of weak ties is virtually zero and that of strong ties is different from zero and positive. The lower panels depict the time-series of the values of the chain, which reveal that stationarity has been achieved.

In terms of magnitude, in an average group of four strong ties, a standard deviation increase in the financial activity of each of the peers translates into a 27% increase of a standard deviation in the individual's financial activity. This yields increases of about 26% in the probability of getting a credit card, 7% in the probability of opening a checking or saving account, 5% in the probability of buying shares, 4% in the probability of getting a loan, and 10% in the probability

[36] When estimating model (3.2) including only strong ties (i.e. $g^W_{ij,r} = 0$), we obtain comparable results.

[37] Observe that we model unobserved factors at the individual level. This means that the unobserved factors affecting weak and strong tie formation may be different.

[38] Borrowing from decision theory, we can say that ϕ^S stochastically dominates ϕ^W, that is $P(\phi^S \geq x) \geq P(\phi^W \geq x), \forall x \in \mathbb{R}$ (first-order stochastic dominance). Figure 3.5 also shows that the distribution of ϕ^S is negatively (left) skewed. This is due to the condition on the autoregressive parameter in spatial models (peer effect parameter) that guarantees matrix inversion in Model (3.2). More specifically, the parameter space is (-0.10,0.10) for our network. While this is never binding for ϕ^W, ϕ^S is constrained to be below the upper bound. See Appendix B for model details.

of having a credit card debt.

3.4.3 Network Formation

For completeness, Table 3.8 reports on the factors driving link formation in Wave I and II. It shows the complete list of estimation results of Model (3.4)-(3.5)-(3.6), that is when network formation and behavior over network are simultaneously estimated.[39] The estimates of the outcome equation (first column) are the ones in column (4) of Table 3.7. Looking at the estimates of the network formation model in the last two columns, one can see that all the significant coefficients are negative. This evidence reveals homophily behavior- the closer two agents are in terms of observable characteristics the higher is the likelihood of a link between them. Interestingly, the factors predicting the existence of a link slightly change between Wave I and Wave II. While family background characteristics (such as parental education and income) are important in Wave I, when the student grows up individual characteristics (such as own income and employment status) acquire more importance. Importantly, it appears that there are unobserved factors which are relevant in network formation both for Wave I and II. Those factors, however, are not correlated with the error term in the outcome equation. Indeed, the estimate of $\sigma_{\varepsilon z}$ is not statistically significant. In our case where networks are quite small, the inclusion of network fixed effects is likely to control for correlated unobservables. As a result, the use of traditional estimation strategies with network fixed effects that treat network formation as exogenous are not likely to produce biased coefficient estimates. This is why our estimates in columns (6) and (7) of Table 3.5 and in columns (3) and (4) in Table 3.7 are similar.

[39] The kernel densities and the time-series of the values of the chain for the parameters of the network formation equation at time t (equation (3.5)) are reported in Appendix D, Figure D4.

3.5 Understanding the mechanism

By exploiting the recent advances in the econometrics of social networks, our estimation strategy accounts for a possible sorting of agents into networks and controls for unobserved individual characteristics. These unobserved factors possibly capture characteristics such as risk aversion and optimism. Having thus ruled out possible effects of confounding factors, we should believe in a causal effect of peers' behavior on individual behavior which depend on the length of the friendship relationship. Thus, the relevant question is why strong ties are important whereas weak ties are not.

One possibility is that when agents have to decide whether to adopt a financial instrument, they face a risk and place higher value on information from (or the behavior of) agents they trust more. Trust has been widely studied as an important driver of financial decisions Guiso et al. (2004, 2008) . The greater the trust in a social tie, the greater the trust in her choice. Repeated interactions play an important role in determining the level of trust. Several theoretical papers explore the role of information transmission and trust formation in communities and networks. Balmaceda and Escobar (2013) model cohesive communities as complete social networks emerging from optimal agents' choices. Agents maximize common knowledge and consequently minimize defection temptation. In their conceptual framework where investors observe whether their direct neighbors invest or not, complete networks are optimal. Their repeated game model with community-based information flow let trust emerge among agents. The repeated interactions horizon generates a bilateral incentive in letting relationships with more trusted agents surviving over time. Karlan et al. (2009) view network connections as a "social collateral" and argue that the level of trust is determined by the structure of the entire network. They focus on borrowing and lending optimal choices in informal contract enforcement by agents joining the network. The utility derived from links prevents agents from acting unfairly and lets them repay the borrowed

value. Kandori (1992) focuses on the role of "social pressure" and "reputation" in informal contracts. Rewarded honesty and punished defection incentivize agents to behave correctly. This incentive is created by repeated interactions among agents.[40] In his model, enforcement mechanisms work best in long-term relationships. Strong correlation patterns in the behavior of connected agents is driven by the presence and circulation of private information among agents.[41,42] Lippert and Spagnolo (2011) explore scenarios characterized by *Word-of-Mouth Communication.* Their game design lets "network closure" be particularly relevant for sustainability of agents relationships, providing a micro-foundation for the idea of "embeddedness" from Granovetter (1985).

The common vein of these theoretical models is broadly that repeated interactions generate trust among agents, who in turn aggregate in cohesive network structures. If our indicator of strong ties captures high level of trust between agents, then an evidence consistent with this line of reasoning should be the finding of an effect of strong ties on individual financial decisions in cohesive network structures only.

Jackson et al. (2012) use the concept of "supported" links to define a "social quilt", i.e. a union of groups of agents where everybody is connected with everybody else (cliques). They provide an analysis of repeated interactions where an individual's decisions are influenced by the network pattern of behavior in the community. Bilateral interactions may not provide natural self-enforcement of cooperation. Any robust equilibrium network must exhibit a specific trait: each of its link (bilateral connection) must be "supported". That is, if some agent i

[40]The Folk Theorem in the repeated game literature (Rubinstein (1979), Fudenberg and Maskin (1986)) provides a formal model of personal enforcement, showing that any mutually beneficial outcome can be sustained as a subgame-perfect equilibrium if the same set of agents frequently play the same stage game ad infinitum.

[41]The role of private information in a community of buyers with word-of-mouth communication is also highlighted by Ahn and Suominen (2001). In this model, buyers receive signals from other agents and adapt their willingness to buy a seller's product. This mechanism incentivizes the seller to produce high quality output.

[42]See also Greif et al. (1994) for an analysis of the role of *bilateral* and *multilateral* reputation mechanisms in the organization of economic transactions.

is linked to an agent j, then there must be some agent k linked to both of them. Agents with "supported" links tend to form tightly knit groups characterized by a relatively high density of ties.[43]

The first panel of Table 3.9 reports the estimation results of Model (3.2) when strong and weak ties are split according to their level of support. The results confirm our conjecture. It indeed appears that the significant correlation between agents' financial decisions arises among strong ties in highly cohesive network structures. Observe that the network structure per se is not a relevant driver of behavior correlation. Indeed, weak ties in highly cohesive networks do not show any similar behavior. It is only when agents have long-lasting friendship relationships that a significant relationship arises. This evidence is thus in line with the idea that a trust-based mechanism is driving our results.

Another possible explanation is that our indicator of strong ties, which is based on the length of the friendship relationship, simply captures the frequency of interactions. This story is not in contrast with our trust-based mechanism described above. Indeed, to the best of our knowledge there is no theoretical model or empirical evidence indicating that the repeated interactions that generate trust should be measured using the length or the frequency of the interactions. However, it is important to understand whether correlated choices of financial products in social networks are to be found only between agents with long lasting friendship ties, or if random, intense encounters in a short amount of time could also be influential. The richness of information provided by the AddHelath allows us to shed light on this issue. More specifically, the Addhealth questionnaire asks detailed questions about the frequency of interactions for each nominated friend. The questions listed are: "Did you go to {NAME}'s house during the past seven

[43] An alternative measure of network connectivity is the clustering coefficient. While clustering is a node-specific measure, support considers pairs of nodes (link-specific measure). Thus, support is more appropriate in our analysis, which is based on bilateral interaction-types (weak or strong). Observe that networks with an high level of clustering will necessarily display a high fraction of supported links, whereas the converse is not true.

days?"; "Did you meet {NAME} after school to hang out or go somewhere during the past seven days?"; "Did you spend time with {NAME} during the past weekend?"; "Did you talk to {NAME} about a problem during the past seven days?"; "Did you talk to {NAME} on the telephone during the past seven days?". We define a *high frequency* friend if the respondent has shared at least two of these activities with the friend, and *low frequency* friend otherwise. The second panel of Table 3.9 shows the estimation results of Model (3.2) when strong and weak ties are split according to the frequency of interactions. It appears that the frequency of interactions does not matters at all. The weak tie effect remains not different from zero, regardless of the strength of interactions, whereas the strong tie effect remains always statistically significant, with no statistical significance in terms of magnitude between high and low frequency strong tie.[44]

3.6 Policy experiments

Using our data and the estimates of the parameters in Model (3.2),[45] we perform Monte Carlo simulations to asses the extent to which the presence of social interactions can alter the effect of exogenous shocks on the financial activity of agents. The simulated shocks are variations in income, which is one of the most important determinants of financial activity. In a simplistic view, an increase in income can be interpreted as a decrease in participation cost, ceteris paribus. Our goal is to provide evidence about the individual and aggregate implications of strong and weak ties effects.

Our analysis can be used to understand which agents (or which type of agents) should be targeted to maximize the aggregate financial activity participation or to converge to a desired distribution of individual financial activity.

Four exercises are implemented. The first three exercises evaluate aggregate

[44] A formal t-test on the difference between high and low frequency strong ties in a pooled model with interaction terms returns a value of 1.45.

[45] The Bayesian estimates in column (4) of Table 3.7 are used.

effects, i.e. the change in the sum of agents' financial activity after a given intervention. In the first exercise the intervention is an increasing income shock for a fixed number of agents (intensive margin) who have a different number of strong ties. In the second, the intervention is a fixed income shock for an increasing number of agents (extensive margin) who have a different number of strong ties. In the third exercise, we increase the income of a fixed number of agents who have no strong ties while decreasing the income of agents who have strong ties and look at the final aggregate financial activity. The fourth exercise reports on individual effects - we increase the income of a given agent while decreasing the income of her/his peers and look at the consequences on her/his individual financial activity.

Figure 3.6 depicts the results for the first two exercises. The surfaces represent the variation of aggregate financial activity in our sample after the simulated shocks. Panel (a) depicts the effect of an increasing positive shock of income (h, x-axis) on aggregate financial activity for agents who have different number of strong ties (n_s, y-axis), holding constant the number of shocked agents. The shock intensity is administered in terms of the estimated standard deviation in our sample (std points). Each point of the surface is an average coming from 500 replications, where in each replication we shock a random sample of nodes of the same numerosity.[46] It appears that the higher is the number of strong ties the shocked agents have, the higher is the aggregate effect of the income shocks. The amplification effects of strong ties is sizable. Indeed, the aggregate effect of an income shock of 10 std points administered to agents that have 4 strong ties is the same as the one of 20 std points administered to agents without strong ties. In panel (b), we increase the number of shocked agents (n_h, y-axis), holding constant

[46]The number of shocked agents is chosen in a way such that for each category of strong ties we use a numerosity not larger than the real one. In our case, the minimum number of agents for each category of strong ties is 13 (when the number of strong ties is equal to 4). We then shock 13 randomly chosen nodes for each category at each replication. The results, however, remain qualitatively unchanged when changing the number of shocked nodes.

the shock intensity.[47] It appears that the aggregate financial activity is higher if the shock is administered to agents with an higher number of strong ties. Indeed, shocking 10 agents who have 4 strong ties produces the same aggregate result as shocking 20 agents who have no strong ties. If policy interventions of this type have a cost, then our results show that targeting highly connected agents can help cutting costs while maintaining the same efficacy. Peer effects can in fact act as a mechanism through which a shock is propagated (and amplified) through the network.

Figure 3.7 shows how the network structure of social ties matters when negative and positive income shocks hit the population. The surface again represents the variation of aggregate financial activity. In this numerical experiment, we increase the income of a fixed number of agents who have no strong ties (i.e. with no network diffusion of their shock),[48] and decrease the income of an increasing number of agents who have a different number of strong ties (i.e. with network diffusion of their shock).[49] We observe that the higher the number of strong ties each shocked agent has, the smaller the number of shocked agents needed to render null the positive shock at the aggregate level. This evidence helps to explain why some policies targeting a large number of agents did not reach the desired effects. Even if the observable costs of using, say, a new digital credit card are lower than the cost incurred when using a traditional product, the social equilibrium may fail to predict the expected rate of adoption of the new credit card. Social interaction effects amplify whatever aggregate local preferences are induced by exogenous cross-product differences in participation costs. Many agents may be discouraged from adopting the new product largely because they do not know anybody else that they trust who has adopted the product. From Figure 3.7

[47] The shock intensity is 2 std points. The results remain qualitatively change when changing the shock intensity.

[48] We set this number equal to 13, as in our previous exercise. The qualitative results, however, do not depend on this number.

[49] The shocks are symmetrical and equal to +2 std points for agents who have no strong ties and equal to -2 std points for those who do have strong ties.

one can see that if highly connected agents have a negative shock, then the aggregate financial activity decreases even if there is a higher number of agents in the economy that experience a positive shock, provided that those agents have lower social interactions. For example, Figure 3.7 reveals that if 11 agents who have 4 strong ties are negatively shocked and 13 agents who have not strong ties are positively shocked, then the aggregate financial activity on average decreases. Social interactions may be responsible for this (seemingly) paradoxical result.

In order to better understand this result, in our last simulation exercise we consider the effects at the individual level of individual and peers' shocks. Each point of the surface represented in Figure 3.8 depicts the variation of individual financial activity after the simulated shocks averaged over 500 replications. In each replication, we randomly extract an individual i who has a certain number of strong ties, increase her/his income by a fixed amount, and decrease each of her/his peer's income by an increasing amount.[50] The exercise is implemented for agents who have a different number of strong ties. We find that the higher the number of strong ties the agent has, the lower the magnitude of the negative shock given to the peers that is needed to cancel the effect of the individual positive shock. For example, Figure 3.8 shows that if an agent has 1 strong tie, then she/he needs the peer's negative shock to be double in absolute value to counterbalance the effect of her/his positive one. However, if the agent has 4 strong ties, it is enough a negative shock equal to one fifth of one's own of each of them .

3.7 Conclusions

In spite of the common consensus about the importance of word-of-mouth on financial product purchases, the finance literature provides little evidence on the

[50]We set the individual income shock equal to 10 std points, while the shock given to the peers varies from -1 to -20 std points. The qualitative results remain qualitatively unchanged when changing such intensities.

role of peer-to-peer communications. Much of the debate is about how to use social media innovatively and effectively. Yet, a large number of consumers rely on offline word-of-mouth when making banking product and brand choices, in particular young customers. The scarcity of studies on face-to face peer effects in finance is mainly motivated by the lack of appropriate data on personal contacts. In addition, endogeneity and reverse causality issues make the identification and estimation of peer effects a challenging empirical exercise.

This paper tries to fill this gap. By employing detailed data on each individual and friends' financial decisions for a representative sample of US students and a novel identification strategy, we are able to uncover the existence and extent of heterogeneous peer effects in financial decisions. Not all social contacts are equally important. Our evidence is consistent with the hypothesis that when agents have to decide whether or not to adopt a financial instrument they face a risk and they might value the information more coming from agents they trust. A social multiplier may amplify consumers' preferences towards certain products. Even if the direct participation costs of adopting, say, a novel digital credit card are lower than the costs incurred with traditional cards, many consumers may be deterred from adopting the new technology largely because they do not know anybody they trust who does so. Thus, if social interaction helps to increase financial market participation, then an effective policy should not only be measured by its direct effects but also by the group interactions it engenders.

Appendix

Appendix A: Descriptive Statistics

Table A1: Data Description and Summary Statistics

Variables	Description	Average (Std.Dev.)	Min - Max
Financial Variables			
Checking Account	Dummy variable taking value one if the respondent has a checking account.	0.76 (0.42)	0 - 1
Saving Account	Dummy variable taking value one if the respondent has a saving account.	0.63 (0.48)	0 - 1
Shares	Dummy variable taking value one if the respondent has any shares of stock in publicly held corporations, mutual funds, or investment trusts, including stocks in IRAs	0.24 (0.43)	0 - 1
Credit Card	Dummy variable taking value one if the respondent has credit card.	0.61 (0.49)	0 - 1
Student Loan	Dummy variable taking value one if the respondent has any student loans or other educational loans that have not yet been paid.	0.33 (0.47)	0 - 1
Credit Card Debt	Dummy variable taking value one if the respondent has any credit card debt.	0.40 (0.49)	0 - 1
Financial Activity Index	The financial activity index is measured using the respondent's financial activities listed above. The index is the first principal component score.	1.47 (0.77)	0 - 2.64
Financial Activity Index of Peers	Sum of financial activity index of respondent's peers.	5.76 (1.81)	0 - 15.10
Individual Socio-demographic Variables			
Male	Dummy variable taking value one if the respondent is male.	0.47 (0.49)	0 - 1
Latino	Race dummies. "White" is the reference group	0.12 (0.33)	0 - 1
Black	//	0.16 (0.37)	0 - 1
Age	Grade of student in the current year.	21.65 (1.58)	18 - 27
Mathematics Score	Mathematics score. Score in mathematics at the most recent grading period, coded as A=4, B=3, C=2, D=1. The variable is zero if missing, a dummy for missing values is included.	2.15 (1.10)	0 - 4
GPA	The school performance is measured using the respondent's scores received in wave II in several subjects, namely English or language arts, history or social science, mathematics, and science. The scores are coded as 1=D or lower, 2=C, 3=B, 4=A. The final composite index is the first principal component score.	1.42 (0.72)	0 - 3.31
Married	Dummy variable taking value one if the respondent is male.	0.16 (0.37)	0 - 1
Family Size	Number of people living in the household	3.36 (1.96)	0 - 10
Employed	Dummy variable taking value one if the respondent is employed.	0.70 (0.46)	0 - 1
Occ. Manager	Occupation dummies. Closest description of the job. Reference category is "other occupation"	0.05 (0.23)	0 - 1
Occ. Prof. Tech.	=	0.17 (0.37)	0 - 1
Occ. Manual	=	0.25 (0.43)	0 - 1
Occ. Sales	=	0.20 (0.38)	0 - 1
Income	Respondent's total yearly personal income before taxes in thousand of dollars, wages or salaries, including tips, bonuses, and overtime pay, and income from self-employment. Interest or dividends (from stocks, bonds, savings, etc.), unemployment insurance, workmen?s compensation, disability, or social security benefits, including SSI (supplemental security income) are included.	14.07 (14.66)	0 - 250

Family Background			
Father Education	Years of education attained by the father of the respondent. The variable is zero if missing. A dummy for missing values is included.	10.73 (6.85)	0 - 19
Parental Income	Total income in thousand of dollars, before taxes of respondent's family. It includes own income, income of everyone else in the household, and income from welfare benefits, dividends, and all other sources.	49.40 (51.42)	0 - 900
Contextual Effects	Average of peers' characteristics of all listed variables.		
Networks			
Links in Wave I	Number of individual links in Wave I.	2.23 (1.88)	0 - 11
Links in Wave II	Number of individual links in Wave II.	2.22 (2.18)	0 - 11
Strong Ties	Percentage of strong ties on total individual links.	0.23 (0.27)	0 - 1
Weak Ties	Percentage of weak ties on total individual links.	0.77 (0.27)	0 - 1

Appendix B: Bayesian Estimation

Prior and Posteriors Distributions

In order to draw random values from the marginal posterior distributions of parameters in Model (3.4)-(3.5)-(3.6) we need to set prior distributions of those parameters. Once priors and likelihoods are specified, we can derive marginal posterior distributions of parameters and draw values from them. Given the link formation Model (3.4)-(3.5), the probability of observing a network r at time t-1 and t, $\mathbf{G}_r^{t-1}$ and $\mathbf{G}_r^t$ is

$$\begin{aligned} P(\mathbf{G}_r^{t-1}|x_{ij,r}, z_{i,r}, z_{j,r}, \gamma_{t-1}, \theta_{t-1}) &= \prod_{i\neq j} P(g_{ij,r,t-1}|x_{ij,r}, z_{i,r}, z_{j,r}, \gamma_{t-1}, \theta_{t-1}), \\ P(\mathbf{G}_r^{t}|x_{ij,r}, z_{i,r}, z_{j,r}, \gamma_{t-1}, \theta_{t-1}) &= \prod_{i\neq j} P(g_{ij,r,t}|x_{ij,r}, z_{i,r}, z_{j,r}, g_{ij,r,t-1}, \gamma_t, \theta_t, \lambda). \end{aligned}$$

Let $\beta^* = (\beta, \delta^S, \delta^W)$, following Hsieh and Lee (2011) our prior distributions are

$$\begin{aligned}
z_{i,r} &\sim N(0,1) \\
\omega &\sim N_{2K+3}(\omega_0, \Omega_0) \\
\phi^S &\sim U[-\kappa_L, \kappa_L] \\
\phi^W &\sim U[-\kappa_S, \kappa_S] \\
\beta^* &\sim N_{3K+1}(\beta_0, B_0) \\
(\sigma_\varepsilon^2, \sigma_{\varepsilon z}) &\sim TN_2(\sigma_0, \Sigma_0) \\
\eta_r|\sigma_\eta &\sim N(0, \sigma_\eta) \\
\sigma_\eta &\sim IG(\frac{\varsigma_0}{2}, \frac{\zeta_0}{2})
\end{aligned}$$

where $\omega = (\gamma_T, \theta_T, \lambda, \gamma_{T-1}, \theta_{T-1})$, $\kappa_L = \frac{1}{\kappa} - |\phi^W|$, $\kappa_S = \frac{1}{\kappa} - |\phi^S|$ and $\kappa = 1/\max(\min(\max_i(\sum_j g_{ij}^S), \max_j(\sum_i g_{ij}^S)), \min(\max_i(\sum_j g_{ij}^W), \max_j(\sum_i g_{ij}^W)))$ from Gershgorin Theorem, $U[\cdot]$, $TN_2(\cdot)$ and $IG(\cdot)$ are respectively the uniform, bivariate truncated normal, and inverse gamma distributions. Those distributions depend on hyper-parameters (like β_0) that are set by the econometrician. It follows that the marginal posteriors are

$$\begin{aligned}
P(\mathbf{Z}_r|\mathbf{Y}_r, \mathbf{G}_r^W, \mathbf{G}_r^S, \rho) &\propto \prod_{r=1}^{\overline{r}} \prod_{i}^{n_r} \phi(z_{i,r}) P(\mathbf{Y}_r, \mathbf{G}_r^W, \mathbf{G}_r^S|\mathbf{Z}_r, \rho) \\
P(\omega|\mathbf{Y}_r, \mathbf{G}_r^W, \mathbf{G}_r^S) &\propto \phi^{2K+3}(\omega, \omega_0, \Omega_0) \prod_{r=1}^{\overline{r}} P(\mathbf{G}_r^W, \mathbf{G}_r^S|\mathbf{Z}_r, \omega) \\
P(\phi^S, \phi^W|\mathbf{Y}_r, \mathbf{G}_r^W, \mathbf{G}_r^S, \mathbf{Z}_r, \beta, \sigma_\varepsilon^2, \sigma_{\varepsilon z}) &\propto \prod_{r=1}^{\overline{r}} P(\mathbf{Y}_r|\mathbf{G}_r^W, \mathbf{G}_r^S, \mathbf{Z}_r, \beta^*, \sigma_\varepsilon^2, \sigma_{\varepsilon z}) \\
P(\beta^*|\mathbf{Y}_r, \mathbf{G}_r^W, \mathbf{G}_r^S, \mathbf{Z}_r, \sigma_\varepsilon^2, \sigma_{\varepsilon z}, \phi^S, \phi^W) &\propto \phi^{3K+2}(\widetilde{\beta}, \widetilde{\mathbf{B}}) \\
P(\sigma_\varepsilon^2, \sigma_{\varepsilon z}|\mathbf{Y}_r, \mathbf{G}_r^W, \mathbf{G}_r^S, \mathbf{Z}_r, \phi^S, \phi^W) &\propto \phi_T^2((\sigma_\varepsilon^2, \sigma_{\varepsilon z}), \sigma_0, \Sigma_0) \prod_{r=1}^{\overline{r}} P(\mathbf{Y}_r|\mathbf{G}_r^W, \mathbf{G}_r^S, \mathbf{Z}_r, \beta^*, \sigma_\varepsilon^2, \sigma_{\varepsilon z}, \sigma_\eta) \\
P(\eta_r|\mathbf{Y}_r, \mathbf{G}_r^W, \mathbf{G}_r^S, \mathbf{Z}_r, \phi^S, \phi^W, \sigma_\varepsilon^2, \sigma_{\varepsilon z}, \sigma_\eta) &\propto \phi(\eta_r, \widetilde{\eta_r}, \widetilde{M_r}) \\
P(\sigma_\eta|\mathbf{Y}_r, \mathbf{G}_r^W, \mathbf{G}_r^S, \mathbf{Z}_r, \phi^S, \phi^W, \sigma_\varepsilon^2, \sigma_{\varepsilon z}) &\propto \iota\gamma(\frac{\varsigma_0 + \overline{r}}{2}, \frac{\zeta_0 + \sum_{r=1}^{\overline{r}} \eta_r^2}{2})
\end{aligned}$$

where $\rho = (\omega, \phi^S, \phi^W, \beta^*, \sigma_\varepsilon^2, \sigma_{\varepsilon z}, \sigma_\eta, \eta)$, $\phi^l(\cdot)$ is the multivariate $l-$ dimensional normal density function, $\phi_T^l(\cdot)$ is the truncated counterpart, $\iota\gamma(\cdot)$ is the inverse gamma density function. $\widetilde{\beta} = \widetilde{B}(B_0^{-1}\beta_0 + \sum_{r=1}^{\overline{r}} \mathbf{X}_r' \mathbf{V}_r (\mathbf{S}_r \mathbf{Y}_r - \sigma_{\varepsilon z} \mathbf{Z}_r))$, $\widetilde{B} = (B_0^{-1} + \sum_{r=1}^{\overline{r}} \mathbf{X}_r' \mathbf{V}_r \mathbf{X}_r)^{-1}$, $\widetilde{\eta_r} = (\sigma_\varepsilon^2 - \sigma_{\varepsilon z}^2)^{-1} \widetilde{M_r} \mathbf{l}_{n_r}' (\mathbf{S}_r \mathbf{Y}_r - \sigma_{\varepsilon z} \mathbf{Z}_r - \mathbf{X}_r^* \beta^*)$,

and $\widetilde{M}_r = (\sigma_\eta^{-2} + (\sigma_\varepsilon^2 - \sigma_{\varepsilon z}^2)^{-1}\mathbf{l}'_{n_r}\mathbf{l}_{n_r})^{-1}$, where $\mathbf{V}_r = (\sigma_\varepsilon^2 - \sigma_{\varepsilon z}^2)I_{n_r} + \sigma_\eta^2\mathbf{l}_{n_r}\mathbf{l}'_{n_r}$, where $\mathbf{X}_r^* = (\mathbf{X}_r, \mathbf{G}_r^{*S}\mathbf{X}_r, \mathbf{G}_r^{*W}\mathbf{X}_r)$. The posteriors of β^*,$\{\eta_r\}$ and σ_η are available in closed forms and a usual Gibbs Sampler is used to draw from them. The other parameters are drawn using the Metropolis-Hastings (M-H) algorithm (Metropolis-within-Gibbs).[51]

Sampling Algorithm

We start our algorithm by picking $(\omega^{(1)}, \phi^{L(1)}, \phi^{S(1)}, \beta^{*(1)}, \sigma_\varepsilon^{2(1)}, \sigma_{\varepsilon z}^{(1)}, \sigma_\eta^{(1)}, \eta^{(1)})$ as starting values. For $\beta^{*(1)}, \eta^{(1)}, \phi^{L(1)}, \phi^{S(1)}$ we use OLS estimates, while we set the variances-covariances $\sigma_\varepsilon^{2(1)}, \sigma_{\varepsilon z}^{(1)}, \sigma_\eta^{(1)}$ at 0.[52] We ought to draw samples of $z_{i,r}^t$ from $P(z_{i,r}|Y_r, G_r^W, G_r^S, \rho),\ i = 1, \cdots, n$. To do this, we first draw a candidate $\widetilde{z}_{i,r}^t$ from a normal distribution with mean $z_{i,r}^{(t-1)}$, then we rely on a M-H decision rule: if $\widetilde{z}_{i,r}^t$ is accepted we set $z_{i,r}^t = \widetilde{z}_{i,r}^t$, otherwise $z_{i,r}^t = z_{i,r}^{t-1}$. Once all $z_{i,r}$ are sampled, we move to the sampling of β^*. By specifying a normal prior and a normal likelihood we can now easily sample β^t from a multivariate normal distribution. A diffuse prior for σ_ϵ^2 allows us to sample it from an inverse chi-squared distribution. We follow the Bayesian spatial econometric literature by sampling ϕ^S, ϕ^W from uniform distributions with support $[-\kappa_L, \kappa_L]$ and $[-\kappa_S, \kappa_S]$, as defined above. A M-H step is then performed over a normal likelihood: if accepted, then $\phi^{S^t} = \widetilde{\phi^S}^t$ and $\phi^{W^t} = \widetilde{\phi^W}^t$. For network fixed effects we deal again with normal prior and normal likelihood, so η is easily sampled from a multivariate normal. We sample $\sigma_\varepsilon^2, \sigma_{\varepsilon z}$ from a truncated bivariate normal over an admissible region Ξ such that the variance-covariance matrix is positive definite. Acceptation or rejection is determined by the usual M-H decision rule. A detailed step-by-step description of the algorithm is provided below.

Step 1: Sample $\mathbf{Z}_r^t$ from $P(\mathbf{Z}_r|\mathbf{Y}_r, \mathbf{G}_r^W, \mathbf{G}_r^S, \rho)$.

[51] See Tierney (1994) and Chib and Greenberg (1996) for details regarding the resulting Markov chain given by the combination of those two methods.

[52] The algorithm is robust to different starting values. However, speed of convergence may increase significantly.

Propose $\widetilde{\mathbf{Z}}_r^t$ drawing each $\widetilde{z}_{i,r}^t$ from $N(z_{i,r}^{(t-1)}, \xi_z)$, then set $z_{i,r}^t = \widetilde{z}_{i,r}^t$ with probability α_Z or $z_{i,r}^t = z_{i,r}^{t-1}$ with probability $1 - \alpha_Z$ where

$$\alpha_Z = \min\left\{ \frac{P(\mathbf{Y}_r|\mathbf{G}_r^W, \mathbf{G}_r^S, \widetilde{\mathbf{Z}}_r^t, \rho^{t-1})}{P(\mathbf{Y}_r|\mathbf{G}_r^W, \mathbf{G}_r^S, \mathbf{Z}_r^{t-1}, \rho^{t-1})} \prod_i^{n_r} \frac{P(g_{ij,r}^W, g_{ij,r}^S|\widetilde{z}_{i,r}^t, z_{j,r}^{t-1}, \omega)}{P(g_{ij,r}^W, g_{ij,r}^S|z_{i,r}^{t-1}, z_{j,r}^{t-1}, \omega)} \frac{\phi(\widetilde{z}_{i,r}^t)}{\phi(z_{i,r}^{t-1})} \right\}$$

Step 2: Sample $\widetilde{\omega}^t$ from $P(\omega|\mathbf{Y}_r, \mathbf{G}_r^W, \mathbf{G}_r^S)$.

Propose $\widetilde{\omega}^t$ from $N^{2K+3}(\omega^{t-1}, \xi_\omega \Omega_0)$, then set $\omega^t = \widetilde{\omega}^t$ with probability α_ω or $\omega^t = \omega^{t-1}$ with probability $1 - \alpha_\omega$ where

$$\alpha_\omega = \min\left\{ \prod_{r=1}^{\overline{r}} \frac{P(\mathbf{G}_r^W, \mathbf{G}_r^S|\mathbf{Z}_r^t, \widetilde{\omega}^t)}{P(\mathbf{G}_r^W, \mathbf{G}_r^S|\mathbf{Z}_r^t, \omega^{t-1})} \frac{\phi^{2K+3}(\widetilde{\omega}^t, \omega_0, \Omega_0)}{\phi^{2K+3}(\omega^{t-1}, \omega_0, \Omega_0)} \right\}$$

Step 3: Sample $\widetilde{\phi^S}^t$ and $\widetilde{\phi^W}^t$ from $P(\phi^S, \phi^W|\mathbf{Y}_r, \mathbf{G}_r^W, \mathbf{G}_r^S, \mathbf{Z}_r, \beta^*, \sigma_\varepsilon^2, \sigma_{\varepsilon z})$.

Propose $\widetilde{\phi^S}^t$ from $N(\phi^{S^{t-1}}, \xi_\phi)$ and $\widetilde{\phi^W}^t$ from $N(\phi^{W^{t-1}}, \xi_\phi)$, then set $\phi^{S^t} = \widetilde{\phi^S}^t$ and $\phi^{W^t} = \widetilde{\phi^W}^t$ with probability α_ϕ or $\phi^{S^t} = \phi^{S^{t-1}}$ and $\phi^{W^t} = \phi^{W^{t-1}}$ with probability $1 - \alpha_\phi$ where

$$\alpha_\phi = \min\left\{ \prod_{r=1}^{\overline{r}} \frac{P(\mathbf{Y}_r|\mathbf{G}_r^W, \mathbf{G}_r^S, \mathbf{Z}_r^{t-1}, \widetilde{\phi^S}^t, \widetilde{\phi^W}^t, \beta^{*t-1}, \sigma_\varepsilon^{2^{t-1}}, \sigma_{\varepsilon z}^{t-1}, \sigma_\eta^{t-1})}{P(\mathbf{Y}_r|\mathbf{G}_r^W, \mathbf{G}_r^S, \mathbf{Z}_r^{t-1}, \phi^{L^{t-1}}, \phi^{S^{t-1}}, \beta^{*t-1}, \sigma_\varepsilon^{2^{t-1}}, \sigma_{\varepsilon z}^{t-1}, \sigma_\eta^{t-1})} \cdot \mathbf{I}(\widetilde{\phi^S}^t \in [-\kappa_L, \kappa_L]) \mathbf{I}(\widetilde{\phi^W}^t \in [-\kappa_S, \kappa_S]) \right\}$$

Step 4: Sample $\widetilde{\sigma}_\varepsilon^t$ and $\widetilde{\sigma}_{\varepsilon z}^t$ from $P(\sigma_\varepsilon^2, \sigma_{\varepsilon z}|\mathbf{Y}_r, \mathbf{G}_r^W, \mathbf{G}_r^S, \mathbf{Z}_r, \phi^S, \phi^W)$.

Propose $\widetilde{\sigma}_\varepsilon^t$ and $\widetilde{\sigma}_{\varepsilon z}^t$ from $N^2((\sigma_\varepsilon^{2^{t-1}}, \sigma_{\varepsilon z}^{t-1}), \xi_\sigma, \Sigma_0)$, then set $\sigma_\varepsilon^t = \widetilde{\sigma}_\varepsilon^t$ and $\sigma_{\varepsilon z}^t = \widetilde{\sigma}_{\varepsilon z}^t$ with probability α_σ or $\sigma_\varepsilon^t = \sigma_\varepsilon^{t-1}$ and $\sigma_{\varepsilon z}^t = \sigma_{\varepsilon z}^{t-1}$ with probability $1 - \alpha_\sigma$ where

$$\alpha_\sigma = \min\left\{ \prod_{r=1}^{\overline{r}} \frac{P(\mathbf{Y}_r|\mathbf{G}_r^W, \mathbf{G}_r^S, \mathbf{Z}_r^{t-1}, \phi^{L^{t-1}}, \phi^{S^{t-1}}, \beta^{*t-1}, \widetilde{\sigma}_\varepsilon^t, \widetilde{\sigma}_{\varepsilon z}^t, \sigma_\eta^{t-1})}{P(\mathbf{Y}_r|\mathbf{G}_r^W, \mathbf{G}_r^S, \mathbf{Z}_r^{t-1}, \phi^{L^{t-1}}, \phi^{S^{t-1}}, \beta^{*t-1}, \sigma_\varepsilon^{t-1}, \sigma_{\varepsilon z}^{t-1}, \sigma_\eta^{t-1})} \frac{\phi_T^2((\widetilde{\sigma}_\varepsilon^t, \widetilde{\sigma}_{\varepsilon z}^t), \sigma_0, \Sigma_0)}{\phi_T^2(\sigma_\varepsilon^{t-1}, \sigma_{\varepsilon z}^{t-1}, \sigma_0, \Sigma_0)} \mathbf{I}((\widetilde{\sigma}_\varepsilon^t, \widetilde{\sigma}_{\varepsilon z}^t) \in \Xi) \right\}$$

where Ξ is a region in which the variance-covariance matrix is definite properly.

Step 5: Sample β^{*t-1}, η^t and σ_η^t from conditional posterior distributions.

Step 6: Repeat previous steps updating values indexed with t.

In each of the M-H steps (1-4) the algorithm accepts the new random values (proposals) if the likelihood is higher than the current one. In the algorithm,

ξ_z ,ξ_ω, ξ_σ, and ξ_ϕ are tuning parameters chosen by the econometrician. This choice determines the rejection rate of proposals in the M-H steps (1-4). We set a dynamic algorithm for calibrating those tuning parameters so that they converge to the optimal ones. Optimality means that the proposals are accepted about 50% of the times.[53] Figure B1 shows the time-series of rejection rates for all of the parameters. It appears that convergence is achieved around an acceptance rate of 50% for all of the parameters.[54]

[53]The intuition is that if a tuning parameter is too high, the draws are less likely to be within "high density regions" of the posterior and then rejection is too frequent. The "step" is too long and the chain "does not move enough". On the other hand if the "step" is too short, the proposal is more likely to be accepted and the chain "moves too much". Given that we want a mixing chain with a balanced proportion of rejections and acceptances, an optimal step must be chosen. Setting it manually requires a huge amount of time and many manual operations. The dynamic setting of tuning parameters is as follows:
if $t_A/t \leq 0.4$ then $\xi_{t+1} = \xi_t/1.1$,
if $t_A/t \geq 0.6$ then $\xi_{t+1} = \xi_t \times 1.1$,
if $0.4 \leq t_A/t \leq 0.6$ then $\xi_{t+1} = \xi_t$,
where t_A is the acceptance rate at iteration t. The procedure decreases the tuning parameter (the "step") when proposals are rejected too frequently, while it increases the tuning parameter when proposals are accepted too frequently. This mechanism guarantees a bounded acceptance rate and convergence to optimal tuning.

[54]Given that the rejection rate-based correction of tuning parameters has 0.4 and 0.6 as boundaries, rejection rates oscillate between these values. The likelihood of reaching the boundaries decreases as the number of draws increases and the rejection rates tend to 0.5, as Figure B1 shows.

Appendix C: IV Estimation

Let $\mathbf{Y}_r = (y_{1,r}, \cdots, y_{n_r,r})'$, $\mathbf{X}_r = (x_{1,r}, \cdots, x_{n_r,r})'$, and $\epsilon_r = (\epsilon_{1,r}, \cdots, \epsilon_{n_r,r})'$. Denote the $n_r \times n_r$ adjacency matrix by $\mathbf{G}_r = [g_{ij,r}]$, the row-normalized of $\mathbf{G}_r$ by $\mathbf{G}_r^*$, and the n_r-dimensional vector of ones by $\mathbf{l}_{n_r}$. Let us split the adjacency matrix into two submatrices $\mathbf{G}_r^S$ and $\mathbf{G}_r^W$, which keep trace of strong and weak ties, respectively. Then, model (3.2) can be written in matrix form as

$$\mathbf{Y}_r = \phi^S \mathbf{G}_r^S \mathbf{Y}_r + \phi^W \mathbf{G}_r^W \mathbf{Y}_r + \mathbf{X}_r^* \beta^* + \eta_r \mathbf{l}_{n_r} + \epsilon_r, \tag{3.10}$$

For a sample with $\bar{r}$ networks, stack up the data by defining $\mathbf{Y} = (\mathbf{Y}_1', \cdots, \mathbf{Y}_{\bar{r}}')'$, $\mathbf{X}^* = (\mathbf{X}_1^{*\prime}, \cdots, \mathbf{X}_{\bar{r}}^{*\prime})'$, $\epsilon = (\epsilon_1', \cdots, \epsilon_{\bar{r}}')'$, $\mathbf{G} = \mathrm{D}(\mathbf{G}_1, \cdots, \mathbf{G}_{\bar{r}})$, $\mathbf{G}^* = \mathrm{D}(\mathbf{G}_1^*, \cdots, \mathbf{G}_{\bar{r}}^*)$, $\iota = \mathrm{D}(\mathbf{l}_{n_1}, \cdots, \mathbf{l}_{n_{\bar{r}}})$ and $\eta = (\eta_1, \cdots, \eta_{\bar{r}})'$, where $\mathrm{D}(\mathbf{A}_1, \cdots, \mathbf{A}_K)$ is a block diagonal matrix in which the diagonal blocks are $n_k \times n_k$ matrices $\mathbf{A}_k$'s. For the entire sample, the model is thus

$$\mathbf{Y} = \phi^S \mathbf{G}^S \mathbf{Y} + \phi^W \mathbf{G}^W \mathbf{Y} + \mathbf{X}^* \beta + \iota \cdot \eta + \epsilon. \tag{3.11}$$

We use the 2SLS estimation strategy from Liu and Lee (2010), and extend it to the case of two different network structures. Model (3.11) can be written as

$$\mathbf{Y} = \mathbf{Z}\theta + \iota \cdot \eta + \epsilon, \tag{3.12}$$

where $\mathbf{Z} = (\mathbf{G^S Y}, \mathbf{G^W Y}, \mathbf{X}^*)$, $\theta = (\phi^S, \phi^W, \beta')'$, and $\iota = \mathrm{D}(\mathbf{l}_{n_1}, \cdots, \mathbf{l}_{n_{\bar{r}}})$.

We treat η as a vector of unknown parameters. When the number of networks $\bar{r}$ is large, we have the incidental parameter problem. Let $\mathbf{J} = \mathrm{D}(\mathbf{J}_1, \cdots, \mathbf{J}_{\bar{r}})$, where $\mathbf{J}_r = \mathbf{I}_{n_r} - \frac{1}{n_r}\mathbf{l}_{n_r}'\mathbf{l}_{n_r}$. The network fixed effect can be eliminated by a transformation with $\mathbf{J}$ such that

$$\mathbf{JY} = \mathbf{JZ}\theta + \mathbf{J}\epsilon. \tag{3.13}$$

Let $\mathbf{M} = (\mathbf{I} - \phi^S \mathbf{G}^S - \phi^W \mathbf{G}^W)^{-1}$. The equilibrium outcome vector $\mathbf{Y}$ in (3.12) is then given by the reduced form equation

$$\mathbf{Y} = \mathbf{M}(\mathbf{X}^* \beta + \iota \cdot \eta) + \mathbf{M}\epsilon. \tag{3.14}$$

It follows that $\mathbf{G}^S\mathbf{Y} = \mathbf{G}^S\mathbf{M}\mathbf{X}^*\beta + \mathbf{G}^S\mathbf{M}\iota\eta + \mathbf{G}^S\mathbf{M}\epsilon$ and $\mathbf{G}^W\mathbf{Y} = \mathbf{G}^W\mathbf{M}\mathbf{X}^*\beta + \mathbf{G}^W\mathbf{M}\iota\eta + \mathbf{G}^W\mathbf{M}\epsilon$. $\mathbf{G}^S\mathbf{Y}$ and $\mathbf{G}^W\mathbf{Y}$ are correlated with ϵ because $\mathrm{E}[(\mathbf{G}^S\mathbf{M}\epsilon)'\epsilon] = \sigma^2\mathrm{tr}(\mathbf{G}^S\mathbf{M}) \neq 0$ and $\mathrm{E}[(\mathbf{G}^W\mathbf{M}\epsilon)'\epsilon] = \sigma^2\mathrm{tr}(\mathbf{G}^W\mathbf{M}) \neq 0$. Hence, in general, (3.13) cannot be consistently estimated by OLS.[55] If $\mathbf{G}$ is row-normalized such that $\mathbf{G} \cdot \mathbf{l}_n = \mathbf{l}_n$, where $\mathbf{l}_n$ is a n-dimensional vector of ones, the endogenous social interaction effect can be interpreted as an average effect.

Liu and Lee (2010) use an instrumental variable approach and propose different estimators based on different instrumental matrices, here denoted by $\mathbf{Q}_1$ and $\mathbf{Q}_2$. In particular, besides the conventional instrumental matrix ($\mathbf{Q}_1 = \mathbf{J}(\mathbf{G}\mathbf{X}^*, \mathbf{X}^*)$) for the estimation of (3.13), they propose to use additional instruments (IVs) $JG\iota$ and enlarge the instrumental matrix $\mathbf{Q}_2 = (\mathbf{Q}_1, JG\iota)$. The additional IVs of $JG\iota$ are simply the row sums of G (i.e. the number of links of each agent). Liu and Lee (2010) show that those additional IVs could help model identification when the conventional IVs are weak and improve on the estimation efficiency of the conventional 2SLS estimator based on $\mathbf{Q}_1$. As a result, an IV based on $\mathbf{Q}_2$ (rather than $\mathbf{Q}_1$) should be preferred. However, the number of such additional instruments depends on the number of networks. If the number of networks grows with the sample size, so does the number of IVs. The 2SLS could be asymptotically biased when the number of IVs increases too quickly relative to the sample size, i.e. when there are many networks. Liu and Lee (2010) thus propose a bias-correction procedure based on the estimated leading-order many-IV bias (*IV bias-corrected*). The bias-corrected IV estimator is properly centered,

[55] Lee (2002) has shown that the OLS estimator can be consistent in the spatial scenario where each spatial unit is influenced by many neighbors whose influences are uniformly small. However, in the current data, the number of neighbors are limited, so that result does not apply.

asymptotically normally distributed, and efficient when the average network size is sufficiently large.[56] The (more efficient) IV estimator (based on $\mathbf{Q}_2$) and its bias-corrected version are the IV estimators used in our analysis.

Let us derive those estimators for equation (3.13), i.e. for the model where agents are heterogeneous and allowed to interact according to different network structures. From the reduced form equation (3.12), we have $E(\mathbf{Z}) = [\mathbf{G}^S\mathbf{M}(\mathbf{X}^*\beta + \iota\cdot\eta), \mathbf{G}^W\mathbf{M}(\mathbf{X}^*\beta+\iota\cdot\eta), \mathbf{X}^*]$. The best IV matrix for $\mathbf{JZ}$ is given by

$$\mathbf{J}f = \mathbf{JE}(\mathbf{Z}) = J[\mathbf{G}^S\mathbf{M}(\mathbf{X}^*\beta+\iota\cdot\eta), \mathbf{G}^W\mathbf{M}(\mathbf{X}^*\beta+\iota\cdot\eta), \mathbf{X}^*] \tag{3.15}$$

which is an $n\times(3m+2)$ matrix. However, this matrix is unfeasible as it involves unknown parameters. Note that f can be considered as a linear combination of the vectors in $\mathbf{Q_0} = J[\mathbf{G}^S\mathbf{M}(\mathbf{X}^*+\iota), \mathbf{G}^W\mathbf{M}(\mathbf{X}^*+\iota), \mathbf{X}^*]$. As ι has $\bar{r}$ columns the number of IVs in $\mathbf{Q_0}$ increases as the number of groups increases. Furthermore, as $\mathbf{M} = (\mathbf{I}-\phi^S\mathbf{G}^S-\phi^W\mathbf{G}^W)^{-1} = \sum_{j=0}^{\infty}(\phi^S\mathbf{G}^S+\phi^W\mathbf{G}^W)^j$ when $\sup||\phi^S\mathbf{G}^S+\phi^W\mathbf{G}^W||_\infty<1$, $\mathbf{MX}^*$ and $\mathbf{M}\iota$ can be approximated by linear combinations of

$$(\mathbf{G}^S\mathbf{X}^*, \mathbf{G}^W\mathbf{X}^*, \mathbf{G}^W\mathbf{G}^S\mathbf{X}^*, \left(\mathbf{G}^S\right)^2\mathbf{X}^*, \left(\mathbf{G}^W\right)^2\mathbf{X}^*, \left(\mathbf{G}^W\right)^2\mathbf{G}^S\mathbf{X}^*, \left(\mathbf{G}^W\right)^2\left(\mathbf{G}^S\right)^2\mathbf{X}^*, \cdots)$$

and

$$(\mathbf{G}^S\iota, \mathbf{G}^W\iota, \mathbf{G}^W\mathbf{G}^S\iota, \left(\mathbf{G}^S\right)^2\iota, \left(\mathbf{G}^W\right)^2\iota, \left(\mathbf{G}^W\right)^2\mathbf{G}^S\iota, \left(\mathbf{G}^W\right)^2\left(\mathbf{G}^S\right)^2\iota, \cdots),$$

respectively. Hence, $\mathbf{Q}_0$ can be approximated by a linear combination of

$$\begin{aligned}\mathbf{Q}_\infty \;=\; & \mathbf{J}(\mathbf{G}^S(\mathbf{G}^S\mathbf{X}^*, \mathbf{G}^W\mathbf{X}^*, \mathbf{G}^W\mathbf{G}^S\mathbf{X}^*, \cdots, \mathbf{G}^S\iota, \mathbf{G}^W\iota, \mathbf{G}^W\mathbf{G}^S\iota, \cdots), \\ & \mathbf{G}^W(\mathbf{G}^S\mathbf{X}^*, \mathbf{G}^W\mathbf{X}^*, \mathbf{G}^W\mathbf{G}^S\mathbf{X}^*, \cdots, \mathbf{G}^S\iota, \mathbf{G}^W\iota, \mathbf{G}^W\mathbf{G}^S\iota, \cdots), \mathbf{X}^*).\end{aligned} \tag{3.16}$$

[56] Liu and Lee (2010) also generalize this 2SLS approach to the GMM using additional quadratic moment conditions.

Let $\mathbf{Q_K}$ be an $n \times K$ submatrix of $\mathbf{Q}_\infty$ (with $K \geq 3m+2$) including $\mathbf{X}^*$. Let $\mathbf{Q}_S$ be an $n \times K_L$ submatrix of $\mathbf{Q}_{L\infty} = \mathbf{G}^S(\mathbf{G}^S\mathbf{X}^*, \mathbf{G}^W\mathbf{X}^*, \mathbf{G}^W\mathbf{G}^S\mathbf{X}^*, \cdots, \mathbf{G}^S\iota, \mathbf{G}^W\iota, \mathbf{G}^W\mathbf{G}^S\iota, \cdots)$ and $\mathbf{Q}_S$ an $n \times K_S$ submatrix of $\mathbf{Q}_{S\infty} = \mathbf{G}^W(\mathbf{G}^S\mathbf{X}^*, \mathbf{G}^W\mathbf{X}^*, \mathbf{G}^W\mathbf{G}^S\mathbf{X}^*, \cdots, \mathbf{G}^S\iota, \mathbf{G}^W\iota, \mathbf{G}^W\mathbf{G}^S\iota, \cdots)$. We assume that $\frac{K_L}{K_S} = 1$. Let $\mathbf{P_K} = \mathbf{Q_K}(\mathbf{Q_K}'\mathbf{Q_K})^{-1}\mathbf{Q_K}'$ be the projector of $\mathbf{Q_K}$. The resulting 2SLS estimator is given by

$$\widehat{\theta}_{2sls} = \left(\mathbf{Z}'\mathbf{P_K}\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{P_K}\mathbf{y}. \tag{3.17}$$

Note that, given that we are in a *multiple adjacency matrices* framework, if the approximation (K_L, K_S) is of high order, the many IV problem can arise even if the number of networks is small. The intuition is the following- the higher the number of adjacency matrices, the higher the number of adjacency matrices' combinations needed for approximating $\mathbf{JE}(\mathbf{Z})$. This should be clear looking at (3.16). If we want to approximate $\mathbf{Q}_\infty$ setting a P-order approximation, we will have $\sum_{p=1}^{P} b^p$ matrices to include, where b is the number of adjacency matrices.

The 2SLS estimators of $\theta = (\phi^S, \phi^W, \beta')'$ considered in this paper are

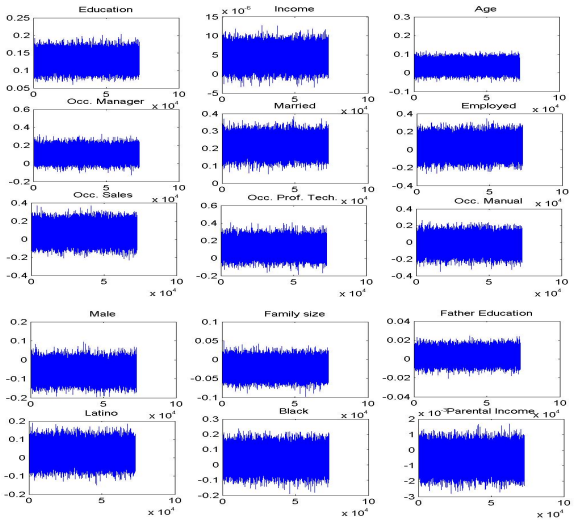
(i) IV : $\widehat{\theta}_{2sls} = (\mathbf{Z}'\mathbf{P}_2\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{P}_2\mathbf{y}$, where $\mathbf{P}_2 = \mathbf{Q}_2(\mathbf{Q}_2'\mathbf{Q}_2)_2^{-1}\mathbf{Q}_2'$ and $\mathbf{Q}_2$ contains the linearly independent columns of $[\mathbf{Q}_1, \mathbf{J}\mathbf{G}^S\iota, \mathbf{J}\mathbf{G}^W\iota]$.

(ii) *IV Bias-corrected*: $\widehat{\theta}_{c2sls} = (\mathbf{Z}'\mathbf{P}_2\mathbf{Z})^{-1}\{\mathbf{Z}'\mathbf{P}_2\boldsymbol{y} - \widetilde{\sigma}^2_{2sls}[\mathrm{tr}\left(\mathbf{P_2}\mathbf{G}^S\widetilde{\mathbf{M}}\right), \mathrm{tr}\left(\mathbf{P_2}\mathbf{G}^W\widetilde{\mathbf{M}}\right), \mathbf{0}_{3m\times 1}]'\}$ where $\tilde{\mathbf{M}} = (\mathbf{I} - \widetilde{\phi}^S_{2sls}\mathbf{G}^S - \widetilde{\phi}^W_{2sls}\mathbf{G}^W)^{-1}$, $\widetilde{\sigma}^2_{2sls}, \widetilde{\phi}^S_{2sls}$ and $\widetilde{\phi}^W_{2sls}$ are $\sqrt{n}$-consistent initial estimators of σ^2, ϕ^S, and ϕ^W obtained by *Finite-IV*. $\widetilde{\sigma}^2_{2sls}[\mathrm{tr}\left(\mathbf{P_2}\mathbf{G}^S\widetilde{\mathbf{M}}\right), \mathrm{tr}\left(\mathbf{P_2}\mathbf{G}^W\widetilde{\mathbf{M}}\right), \mathbf{0}_{3m\times 1}]$ is the empirical counterpart of the theoretical many-IV bias $b_{2sls} = \sigma^2\left(\mathbf{Z}'\mathbf{P_K}\mathbf{Z}\right)^{-1}[\mathrm{tr}\left(\mathbf{\Psi}_{K,L}\right), \mathrm{tr}\left(\mathbf{\Psi}_{K,S}\right), \mathbf{0}'_{3m\times 1}]$, where $\mathbf{\Psi}_{K,L} = \mathbf{P_K}\mathbf{G}^S\mathbf{M}$ and $\mathbf{\Psi}_{K,S} = \mathbf{P_K}\mathbf{G}^W\mathbf{M}$.

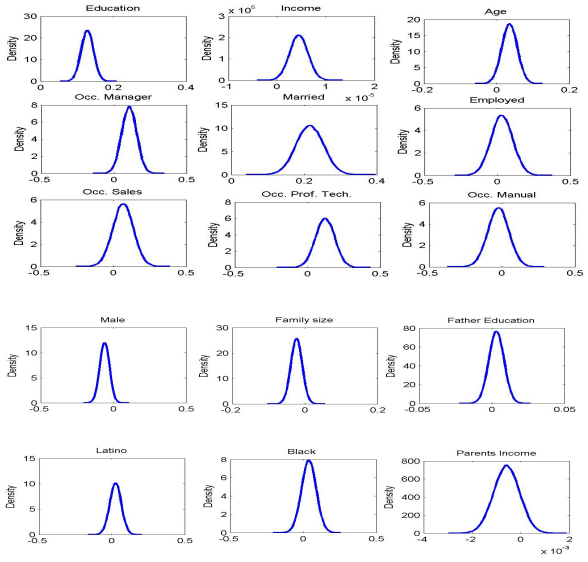
Appendix D: Bayesian Estimation - Additional Results

Figure D1: Bayesian Estimation Results
Control Variables (β)

Panel (a)



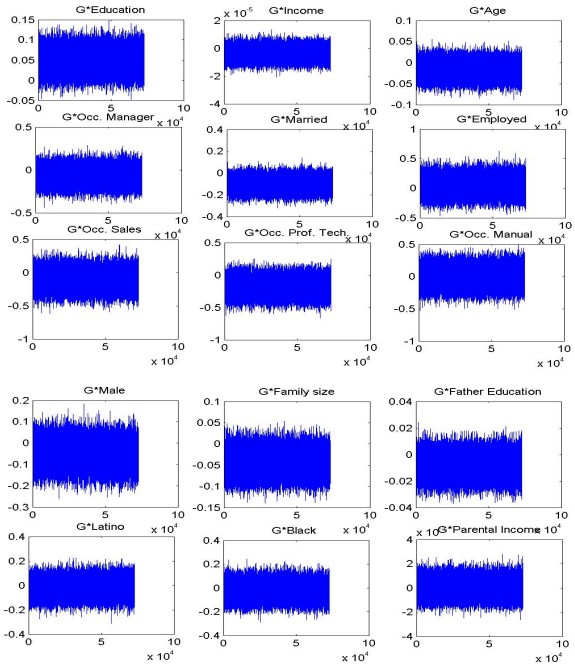
Panel (b)



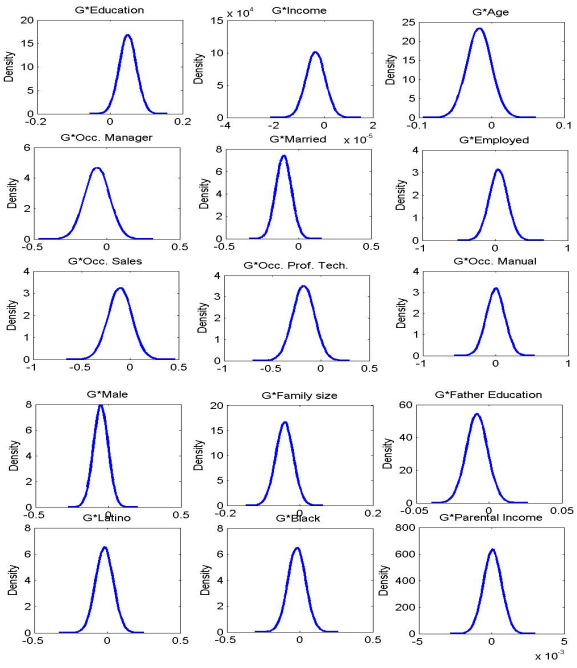
Notes: see Figure 3.3.

Figure D2: Bayesian Estimation Results
Contextual Effects (δ)

Panel (a)



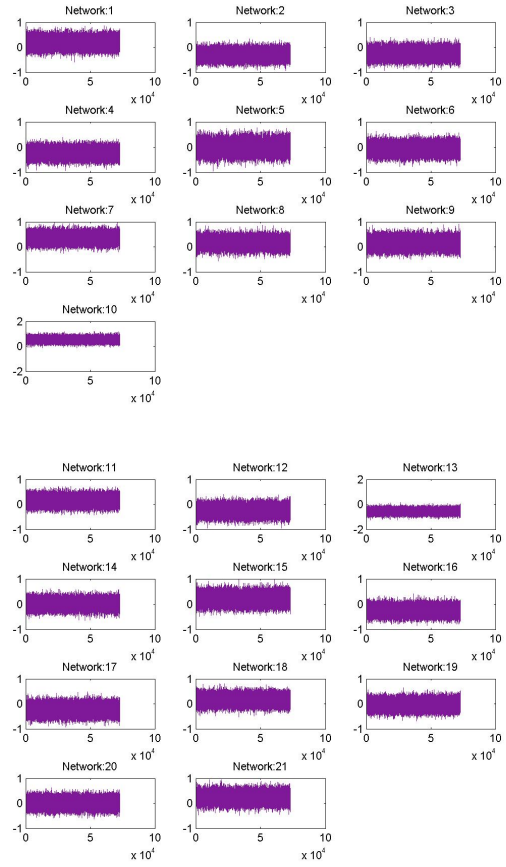
Panel (b)



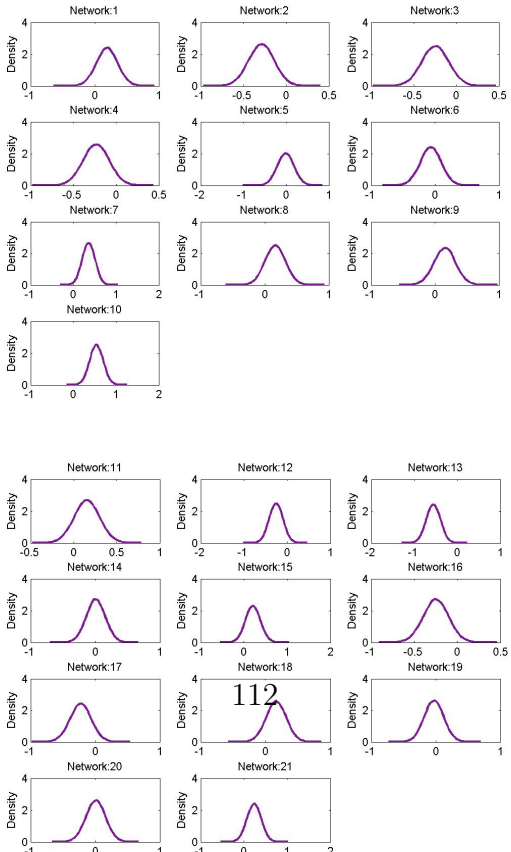
Notes: see Figure 3.3.

Figure D3: Bayesian Estimation Results
Network Fixed Effects (η)

Panel (a)



Panel (b)



Appendix E: Results for the Entire Sample

Table E1: Peer Effects in Financial Decisions
-Entire Sample-

Dependent Variable: Financial Activity Index						
	OLS	OLS	OLS	ML	IV	IV bias-corrected
	(1)	(2)	(3)	(4)	(5)	(6)
Peer Effects ϕ	**0.0520*** **	**0.0450*** **	**-0.0081**	**0.0308*** **	**0.0779*** **	**0.0451** **
	(0.0145)	**(0.0184)**	**(0.0184)**	**(0.0133)**	**(0.0233)**	**(0.0234)**
Male	-0.0950***	-0.0980***	-0.1046***	-0.1027***	-0.1095***	-0.1102***
	(0.0360)	(0.0374)	(0.0366)	(0.0400)	(0.0383)	(0.0381)
Latino	-0.0089	0.0251	0.0342	-0.0059	0.0228	0.0137
	(0.0731)	(0.0796)	(0.0868)	(0.0867)	(0.0908)	(0.0905)
Black	-0.1239***	-0.1267**	0.0419	-0.1706***	0.0694	0.0559
	(0.0466)	(0.0583)	(0.0886)	(0.0648)	(0.0927)	(0.0924)
Age	-0.0094	-0.0051	-0.0172	-0.0956***	-0.0054	-0.0022
	(0.0127)	(0.0140)	(0.0173)	(0.0096)	(0.0178)	(0.0177)
Education	0.1463***	0.1456***	0.1261***	0.1499***	0.1218***	0.1246***
	(0.0116)	(0.0119)	(0.0123)	(0.0125)	(0.0129)	(0.0129)
Income	6.32E-06***	6.21E-06***	5.99E-06***	9.31E-06***	6.00E-06***	6.06E-06***
	(1.45E-06)	(1.47E-06)	(1.47E-06)	(1.81E-06)	(1.54E-06)	(1.53E-06)
Employed	0.2451***	0.2465***	0.2773***	0.2672***	0.2825***	0.2854***
	(0.0685)	(0.0694)	(0.0684)	(0.0749)	(0.0714)	(0.0711)
Occ. Manager	0.2112	0.2330	0.2367	0.2513	0.3295**	0.3422**
	(0.1712)	(0.1700)	(0.1817)	(0.1872)	(0.1629)	(0.1718)
Occ. Prof. Tech	-0.1247*	-0.1310*	-0.1238	-0.1205	-0.1122	-0.1191
	(0.0756)	(0.0764)	(0.0750)	(0.0807)	(0.0787)	(0.0784)
Occ. Manual	-0.1741***	-0.1864***	-0.1848***	-0.2255***	-0.1818***	-0.1827***
	(0.0690)	(0.0698)	(0.0689)	(0.0750)	(0.0718)	(0.0715)
Occ. Sales	-0.0591	-0.0591	-0.0619	-0.0695	-0.0609	-0.0651
	(0.0725)	(0.0730)	(0.0723)	(0.0777)	(0.0757)	(0.0754)
Married	0.3267***	0.3289***	0.3719***	0.3879***	0.3575***	0.3618***
	(0.0510)	(0.0522)	(0.0519)	(0.0537)	(0.0540)	(0.0538)
Family Size	-0.0247**	-0.0230*	-0.0233*	-0.0346***	-0.0221*	-0.0245*
	(0.0118)	(0.0120)	(0.0120)	(0.0124)	(0.0126)	(0.0125)
Father Education	0.0204**	0.0226***	0.0069	-0.0016	0.0091	0.0100
	(0.0083)	(0.0084)	(0.0089)	(0.0086)	(0.0093)	(0.0093)
Parental Income	0.0001	0.0001	-0.0001	0.0003	-0.0002	-0.0002
	(0.0003)	(0.0003)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Constant	-2.1605***	-2.2904***		-2.3642***		
	(0.3047)	(0.3227)		(0.4822)		
School Performance Variables	Yes	Yes	Yes	Yes	Yes	Yes
Contextual Effects	No	Yes	Yes	Yes	Yes	Yes
Network Fixed Effects	No	No	Yes	No	Yes	Yes
Number of Observations	1497	1497	1497	1497	1497	1497
Number of Networks	151	151	151	151	151	151

Notes: standard errors in parentheses. *** p<0.01,** p<0.05,* p<0.1. Dummy variables for missing Income, Family Size, Father Education, Parental Income and GPA are included. Maximum network size 400, minimum 4.

Table E2: Weak and Strong Ties in Financial Decisions
-Entire Sample-

Dependent Variable: Financial Activity Index			
	OLS	IV	IV bias-corrected
	(1)	(2)	(3)
Strong Ties ϕ^S	**0.0526****	**0.1571*****	**0.0443****
	(0.0215)	**(0.0221)**	**(0.0221)**
Weak Ties ϕ^W	**0.0228**	**-0.0700**	**0.0237**
	(0.0169)	**(0.0425)**	**(0.0427)**
Male	-0.0965***	-0.1005***	-0.0962***
	(0.0383)	(0.0399)	(0.0415)
Latino	0.0407	0.0645	0.0817
	(0.0812)	(0.0904)	(0.0940)
Black	-0.1557***	0.0771	0.0882
	(0.0657)	(0.0947)	(0.0985)
Age	-0.0114	-0.0266	-0.0342*
	(0.0146)	(0.0190)	(0.0197)
Education	0.1431***	0.1227***	0.1192***
	(0.0121)	(0.0130)	(0.0135)
Income	0.0000***	0.0000***	0.0000***
	(0.0000)	(0.0000)	(0.0000)
Employed	0.2432***	0.3045***	0.3115***
	(0.0696)	(0.0722)	(0.0751)
Occ. Manager	0.2493	0.3762**	0.3482**
	(0.1703)	(0.1681)	(0.1711)
Occ. Prof. Tech	-0.1337*	-0.1363*	-0.1342
	(0.0770)	(0.0798)	(0.0830)
Occ. Manual	-0.1689***	-0.1958***	-0.2041***
	(0.0701)	(0.0728)	(0.0757)
Occ. Sales	-0.0447	-0.0844	-0.0950
	(0.0733)	(0.0771)	(0.0801)
Married	0.3493***	0.3982***	0.4052***
	(0.0526)	(0.0553)	(0.0575)
Family Size	-0.0236*	-0.0267**	-0.0268**
	(0.0121)	(0.0128)	(0.0133)
Father Education	0.0178**	0.0044	0.0027
	(0.0086)	(0.0095)	(0.0099)
Parental Income	0.0001	-0.0001	0.0000
	(0.0003)	(0.0004)	(0.0004)
Constant	-2.0534***		
	0.3384		
School Performance Variables	Yes	Yes	Yes
Contextual Effects	Yes	Yes	Yes
Network Fixed Effects	No	Yes	Yes
Number of Observations	1497	1497	1497
Number of Networks	151	151	151

Notes: see Table E1.

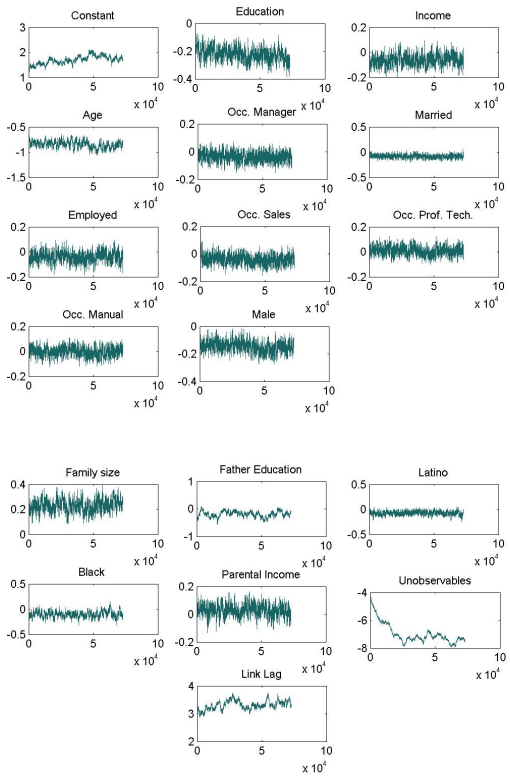
Table 3.4: Financial Activity Participation

	Percentage of Agents Possessing	Contribution to the Financial Activity Index
Checking Account	76%	0.40
Credit Card	61%	0.57
Saving Account	63%	0.73
Shares	25%	0.80
Student Loan	33%	0.53
Credit Card Debt	41%	0.47

Notes: the Financial Activity Index is obtained using a principal component analysis on the listed variables. It is the first principal component, which explains 35 % of the total variance.

Figure D4: Bayesian Estimation Results
Link Formation Control Variables (ω)

Panel (a)



Panel b)

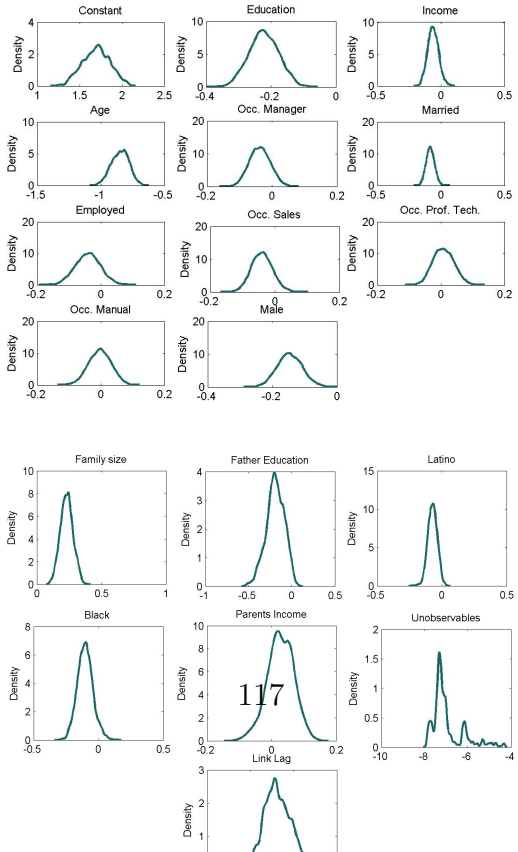


Table 3.5: Peer Effects in Financial Decisions

Dependent Variable: Financial Activity Index

	OLS	OLS	OLS	ML	IV	IV bias-corrected	Bayesian
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Peer Effects(ϕ)	**0.0783*** **	**0.0861*** **	**0.0696*** **	**0.0524*** **	**0.0873*** **	**0.0538***	**0.0518*** **
	(0.0215)	**(0.0282)**	**(0.0279)**	**(0.0189)**	**(0.0322)**	**(0.0322)**	**(0.0162)**
Male	-0.0952*	-0.1009*	-0.0989*	-0.1155*	-0.1110*	-0.1095*	-0.0605*
	(0.0585)	(0.0615)	(0.0594)	(0.0668)	(0.0582)	(0.0582)	(0.0330)
Latino	0.0895	0.1502	0.1771	0.0842	0.1686	0.1644	0.0254
	(0.1241)	(0.1318)	(0.1294)	(0.1441)	(0.1312)	(0.1312)	(0.0393)
Black	-0.1486	-0.1913	0.1789	-0.2338	0.2193	0.2210	0.0385
	(0.1052)	(0.1325)	(0.1519)	(0.1445)	(0.1522)	(0.1522)	(0.0502)
Age	0.0068	0.0074	0.0122	-0.0758***	0.0323	0.0342	0.0358*
	(0.0205)	(0.0232)	(0.0256)	(0.0163)	(0.0248)	(0.0248)	(0.0214)
Education	0.1192***	0.1204***	0.1010***	0.1116***	0.0919***	0.0947***	0.1286***
	(0.0196)	(0.0202)	(0.0199)	(0.0217)	(0.0198)	(0.0198)	(0.0170)
Income	3.72E-06*	4.31E-06**	4.04E-06**	8.74E-06***	4.01E-06**	4.13E-06**	4.44E-06***
	(2.10E-06)	(2.13E-06)	(2.07E-06)	(3.53E-06)	(2.04E-06)	(2.03E-06)	(1.87E-06)
Employed	0.0729	0.1247	0.1118	0.1163	0.0517	0.0460	0.0310
	(0.1487)	(0.1511)	(0.1465)	(0.1603)	(0.1433)	(0.1433)	(0.0745)
Occ. Manager	0.2322	0.2430	0.2378	0.2616	0.3355**	0.3407**	0.1056**
	(0.1802)	(0.1830)	(0.1757)	(0.1942)	(0.1719)	(0.1718)	(0.0510)
Occ. Prof. Tech.	0.1408	0.1107	0.1146	0.1136	0.1804	0.1787	0.1217*
	(0.1570)	(0.1592)	(0.1537)	(0.1690)	(0.1509)	(0.1508)	(0.0658)
Occ. Manual	0.0025	-0.0548	-0.0652	-0.1013	0.0117	0.0204	-0.0243
	(0.1488)	(0.1514)	(0.1465)	(0.1611)	(0.1437)	(0.1436)	(0.0719)
Occ. Sales	0.0830	0.0529	0.0460	0.0652	0.1058	0.1118	0.0679
	(0.1521)	(0.1543)	(0.1511)	(0.1634)	(0.1484)	(0.1483)	(0.0705)
Married	0.3018***	0.3112***	0.3618***	0.3353***	0.3521***	0.3521***	0.2159***
	(0.0778)	(0.0799)	(0.0792)	(0.0843)	(0.0779)	(0.0779)	(0.0375)

Family Size	-0.0147	-0.0152	-0.0170	-0.0305	-0.0088	-0.0126	-0.0229
	(0.0192)	(0.0196)	(0.0190)	(0.0204)	(0.0188)	(0.0188)	(0.0155)
Father Education	0.0215*	0.0272**	0.0028	0.0101	0.0053	0.0068	0.0032
	(0.0132)	(0.0136)	(0.0143)	(0.0140)	(0.0140)	(0.0140)	(0.0052)
Parental Income	0.0006	0.0004	-0.0006	0.0003	-0.0006	-0.0006	-0.0006
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0005)
Constant	-2.1339***	-2.2749***		-2.3442***			
	(0.4887)	(0.5312)		(0.4912)			
$\sigma_{\epsilon \mathbf{z}}$							**-0.0860**
							(0.0534)
School Performance Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Contextual Effects	No	Yes	Yes	Yes	Yes	Yes	Yes
Network Fixed Effects	No	No	Yes	No	Yes	Yes	Yes
Number of Observations	569	569	569	569	569	569	569
Number of Networks	21	21	21	21	21	21	21

Notes: standard errors in parentheses. *** p<0.01,** p<0.05,* p<0.1. Dummy variables for missing Income, Family Size, Father Education, Parental Income and GPA are included. Maximum network size 50, minimum 10.

Table 3.6: 2SLS First Stage Results

Dependent Variable: GY			
	X	GX	G^2X
Variables: X	Own	Peers	Peers of peers (Exclusion Restrictions)
Male	-0.1113	-0.0694	0.0159
	(0.0830)	(0.0655)	(0.0341)
Latino	0.3363**	0.1356	-0.4939***
	(0.1731)	(0.1441)	(0.0674)
Black	-0.0420	0.4622***	0.1238
	(0.2032)	(0.1874)	(0.0968)
Age	-0.0294	0.0143	0.0231***
	(0.0350)	(0.0282)	(0.0078)
Education	0.0700***	0.1323***	-0.0031
	(0.0277)	(0.0223)	(0.0111)
Income	3.61E-07	8.24E-06***	-8.19E-07
	(2.68E-06)	(3.50E-06)	(1.69E-06)
Employed	0.2556	-0.1312	-0.2035***
	(0.2114)	(0.1583)	(0.0842)
Occ. Manager	0.0877	0.3160	-0.1892
	(0.2548)	(0.2038)	(0.1171)
Occ. Prof. Tech.	-0.1222	0.1921	-0.0868
	(0.2256)	(0.1682)	(0.0922)
Occ. Manual	-0.0747	0.0374	0.0247
	(0.2102)	(0.1639)	(0.0848)
Occ. Sales	-0.1370	0.1682	0.0263
	(0.2171)	(0.1671)	(0.0863)
Married	-0.2654***	0.4318***	0.1468***
	(0.1168)	(0.0796)	(0.0444)
Family Size	0.0359	-0.0389*	-0.0423***
	(0.0266)	(0.0211)	(0.0116)
Father Education	0.0317	-0.0065	-0.0051
	(0.0202)	(0.0146)	(0.0073)
Parental Income	0.0006	-0.0021***	-0.0004
	(0.0010)	(0.0006)	(0.0004)
F-stat			**10.8892**
School Performance Variables		Yes	
Network Fixed Effects		Yes	
Number of Observations		Yes	
Number of Networks		Yes	

Notes: OLS estimation results, standard errors in parentheses.*** p<0.01,** p<0.05,* p<0.1. Dummy variables for missing values in variables are included, see Table 3.5. The instrumental set also includes the individual number of connections. See Appendix C for further details on IV estimation of spatial models.

Table 3.7: Weak and Strong Ties in Financial Decisions

Dependent Variable: Financial Activity Index

	OLS	IV	IV bias-corrected	Bayesian
	(1)	(2)	(3)	(4)
Strong Ties (ϕ^S)	**0.1686*** **	**0.1755*** **	**0.0671***	**0.0707*** **
	(0.0318)	**(0.0398)**	**(0.0402)**	**(0.0158)**
Weak Ties (ϕ^W)	**-0.0335**	**-0.0295**	**0.0128**	**-0.0027**
	(0.0286)	**(0.0206)**	**(0.0208)**	**(0.0123)**
Male	-0.0667	-0.0720	-0.0670	-0.0257
	(0.0634)	(0.0604)	(0.0603)	(0.0340)
Latino	0.1309	0.1356	0.1580	0.0138
	(0.1348)	(0.1297)	(0.1295)	(0.0392)
Black	-0.1097	0.2584*	0.3204**	0.0653
	(0.1405)	(0.1573)	(0.1570)	(0.0540)
Age	0.0058	0.0104	0.0165	0.0205
	(0.0245)	(0.0262)	(0.0262)	(0.0220)
Education	0.1097***	0.0913***	0.0996***	0.1327***
	(0.0205)	(0.0198)	(0.0198)	(0.0169)
Income	3.13E-06	2.63E-06	2.91E-06	3.54E-06**
	(2.13E-06)	(2.00E-06)	(1.99E-06)	(1.82E-06)
Employed	0.0711	-0.0313	-0.0131	0.0334
	(0.1530)	(0.1447)	(0.1445)	(0.0752)
Occ. Manager	0.2653	0.3662**	0.3461**	0.0935*
	(0.1870)	(0.1761)	(0.1757)	(0.0527)
Occ. Prof. Tech.	0.2151	0.2689*	0.2363	0.1332**
	(0.1638)	(0.1558)	(0.1556)	(0.0678)
Occ. Manual	0.0139	0.0783	0.0731	-0.0099
	(0.1534)	(0.1452)	(0.1449)	(0.0729)
Occ. Sales	0.1197	0.1755	0.1762	0.0760
	(0.1560)	(0.1501)	(0.1498)	(0.0717)
Married	0.3849***	0.4404***	0.4097***	0.2248***
	(0.0844)	(0.0807)	(0.0806)	(0.0389)
Family Size	-0.0063	-0.0116	-0.0143	-0.0200
	(0.0199)	(0.0189)	(0.0188)	(0.0153)
Father Education	0.0229*	0.0016	0.0062	0.0021
	(0.0140)	(0.0141)	(0.0141)	(0.0050)
Parental Income	0.0005	-0.0002	-0.0006	-0.0004
	(0.0006)	(0.0006)	(0.0006)	(0.0005)
Constant	-2.0713***			
	0.5558			
$\sigma_{\epsilon z}$				**-0.0338**
				(0.0643)
School Performance Variables	Yes	Yes	Yes	Yes
Contextual Effects	Yes	Yes	Yes	Yes
Network Fixed Effects	No	Yes	Yes	Yes
Number of Observations	569	569	569	569
Number of Networks	21	21	21	21

Notes: see Table 3.5.

Table 3.8: Network Formation and Financial Activity
Bayesian Estimation

	Outcome	Link Formation	
(l10ptr10pt)3-4		$t-1$	t
Strong Ties	0.0707***		
	(0.0158)		
Weak Ties	-0.0027		
	(0.0123)		
Male	-0.0257	-0.0831***	-0.1667***
	(0.0340)	(0.0212)	(0.0244)
Age	0.0205	-1.0166***	-1.1772***
	(0.0220)	(0.0604)	(0.0820)
Latino	0.0138	-0.0579***	-0.1441***
	(0.0392)	(0.0215)	(0.0296)
Black	0.0653	-0.1783***	-0.2340***
	(0.0540)	(0.0408)	(0.0578)
Education	0.1327***	-0.1493***	-0.1968***
	(0.0169)	(0.0266)	(0.0317)
Income	3.54E-06**	-0.0487	-0.1770***
	(1.82E-06)	(0.0308)	(0.0386)
Employed	0.0334	-0.0290	-0.0647***
	(0.0752)	(0.0214)	(0.0242)
Occ. Manager	0.0935*	-0.0069	-0.0418*
	(0.0527)	(0.0180)	(0.0237)
Occ. Prof. Tech.	0.1332**	-0.0367*	0.0376
	(0.0678)	(0.0203)	(0.0242)
Occ. Manual	-0.0099	-0.0641***	-0.0560**
	(0.0729)	(0.0220)	(0.0259)
Occ. Sales	0.0760	-0.0580***	-0.0051
	(0.0717)	(0.0183)	(0.0241)
Married	0.2248***	0.0043	0.0008
	(0.0389)	(0.0195)	(0.0237)
Family Size	-0.0200	0.0391	0.0548
	(0.0153)	(0.0255)	(0.0343)
Father Education	0.0021	-0.1797***	-0.0658
	(0.0050)	(0.0587)	(0.0910)
Parental Income	-0.0004	-0.0379	-0.0453
	(0.0005)	(0.0294)	(0.0354)
Constant		-0.7269***	-1.2700***
		(0.0712)	(0.1028)
Link at t-1 ($g_{ij,t-1}$)			1.4096***
			(0.0704)
Unobservables (z)		0.6891***	0.9642***
		(0.0549)	(0.0698)
$\sigma_{\epsilon z}$	-0.0338		
	(0.0643)		
σ_{ϵ}	0.7062		
	(0.3235)		

School Performance Variables	Yes	Yes	Yes
Contextual Effects	Yes	Yes	Yes
Network Fixed Effects	Yes	Yes	Yes
Number of Observations	569	18985	18985
Number of Networks	21	21	21

Notes: see Table 3.5. We report peer effects estimate when network formation an behavior over network are jointly considered. Column (1) reports on the results for Model (2), columns (2)-(3) report on the results for Model (4)-(5).

Table 3.9: Understanding the Mechanism

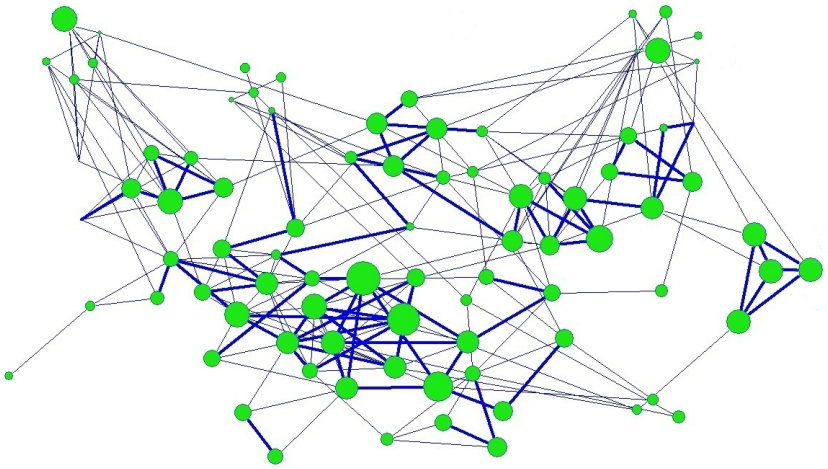
Dependent Variable: Financial Activity Index

(l10ptr10pt)2-4 (l10ptr10pt)5-7	Network Topology			Frequency of Interactions		
	% of links	OLS	IV bias-corrected	% of links	OLS	IV bias-corrected
Strong Ties Supported	72%	0.1646***	0.0691**			
		(0.0316)	(0.0318)			
Strong Ties not Supported	28%	0.1923***	0.0409			
		(0.0347)	(0.0357)			
Weak Ties Supported	58%	-0.0037	0.0266			
		(0.0276)	(0.0274)			
Weak Ties not Supported	42%	-0.0575	-0.0053			
		(0.0318)	(0.0314)			
Strong Ties High Frequency				78%	0.1589***	0.0675**
					(0.0223)	(0.0218)
Strong Ties Low Frequency				22%	0.2543***	0.0790**
					(0.0357)	(0.0367)
Weak Ties High Frequency				84%	-0.0272	-0.0114
					(0.0273)	(0.0276)
Weak Ties Low Frequency				16%	-0.0318	0.0073
					(0.0549)	(0.0559)
Male		-0.0657	-0.0703		-0.0699	-0.0650
		(0.0610)	(0.0604)		(0.0610)	(0.0606)
Latino		0.1756	0.1876		0.1708	0.1837
		(0.1313)	(0.1293)		(0.1315)	(0.1298)
Black		0.2567*	0.3269**		0.2504	0.3286**
		(0.1548)	(0.1575)		(0.1540)	(0.1564)
Age		0.0102	0.0146		0.0093	0.0113
		(0.0266)	(0.0261)		(0.0266)	(0.0263)
Education		0.0935***	0.0991***		0.0931***	0.1005***
		(0.0201)	(0.0197)		(0.0201)	(0.0197)
Income		0.0000	0.0000		0.0000	0.0000
		(0.0000)	(0.0000)		(0.0000)	(0.0000)
Employed		0.0210	0.0061		0.0075	-0.0102
		(0.1476)	(0.1449)		(0.1475)	(0.1453)
Occ. Manager		0.3036*	0.3164*		0.3193*	0.3359**
		(0.1786)	(0.1757)		(0.1787)	(0.1764)
Occ. Prof. Tech		0.2472	0.2187		0.2501	0.2187
		(0.1571)	(0.1555)		(0.1567)	(0.1557)
Occ. Manual		0.0408	0.0589		0.0483	0.0672
		(0.1478)	(0.1455)		(0.1474)	(0.1455)
Occ. Sales		0.1546	0.1633		0.1624	0.1727
		(0.1515)	(0.1504)		(0.1514)	(0.1507)
Married		0.4169***	0.3943***		0.4211***	0.4099***
		(0.0830)	(0.0797)		(0.0829)	(0.0800)
Family Size		-0.0088	-0.0116		-0.0104	-0.0159
		(0.0193)	(0.0188)		(0.0194)	(0.0189)
Father Education		0.0042	0.0058		0.0051	0.0044
		(0.0144)	(0.0140)		(0.0145)	(0.0141)

Parental Income	-0.0002	-0.0007	-0.0002	-0.0007
	(0.0006)	(0.0006)	(0.0006)	(0.0006)
School Performance Variables	Yes	Yes	Yes	Yes
Contextual Effects	Yes	Yes	Yes	Yes
Network Fixed Effects	Yes	Yes	Yes	Yes
Number of Observations	569	569	569	569
Number of Networks	21	21	21	21

Notes: see Table 3.5. Percentage of links is referred to the total of same type of tie (strong or weak).

Figure 3.1: Social Ties and Financial Activity



Notes: a network of 49 agents (nodes) is represented. The size of the node is proportional to the agent's financial activity; the thickness of lines is proportional to the length of the relationship between agents. Thicker lines represent strong ties, while thinner ones represent weak ties.

Figure 3.2: Identification with Network Data

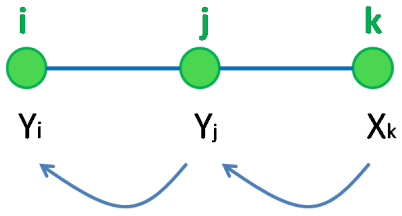
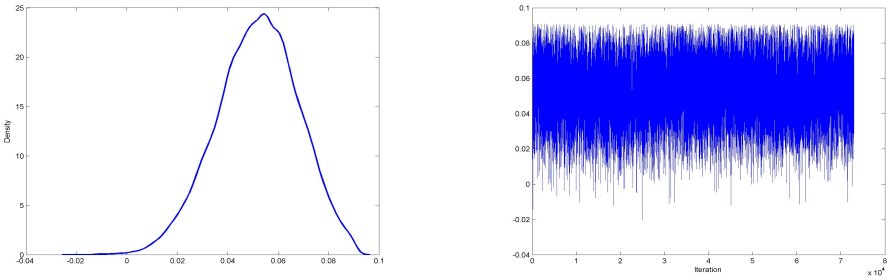


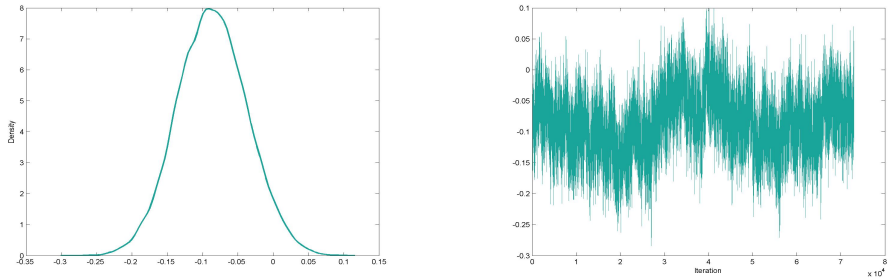
Figure 3.3: Bayesian Estimation Results
Peer Effects (ϕ)



(a) Posterior Distribution (b) Markov Chain

Notes: panel (a) shows the kernel density estimate of the posterior distribution. Panel (b) shows the Markov chain draws.

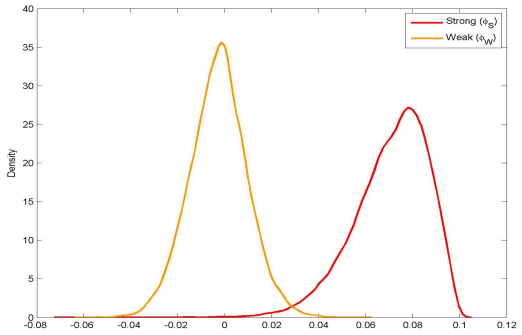
Figure 3.4: Bayesian Estimation Results
Covariance between Unobservables ($\sigma_{\epsilon,z}$)



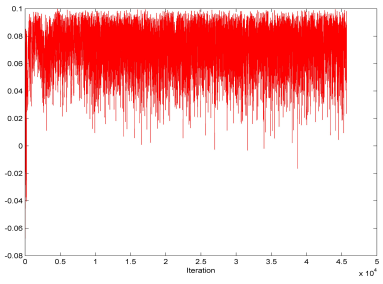
(a) Posterior Distribution (b) Markov Chain

Notes: panel (a) shows the kernel density estimate of the posterior distribution. Panel (b) shows the Markov chain draws.

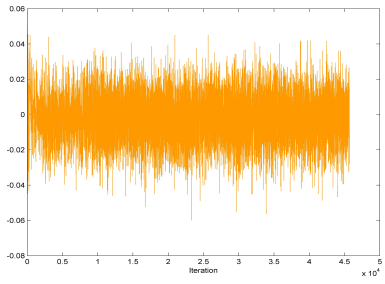
Figure 3.5: Bayesian Estimation Results.
Strong (ϕ_S) vs Weak (ϕ_W) Tie Effects



(a) Posterior Distributions



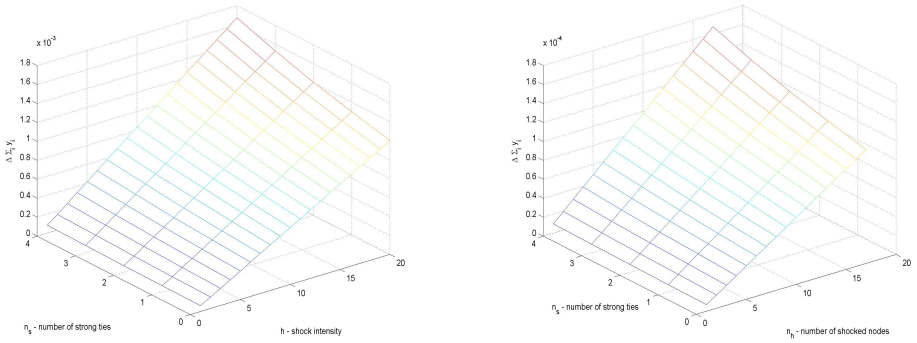
(b) Markov Chain



(c) Markov Chain

Notes: panel (a) shows the kernel density estimates of the posterior distributions. Panel (b) and panel (c) show the Markov chain draws.

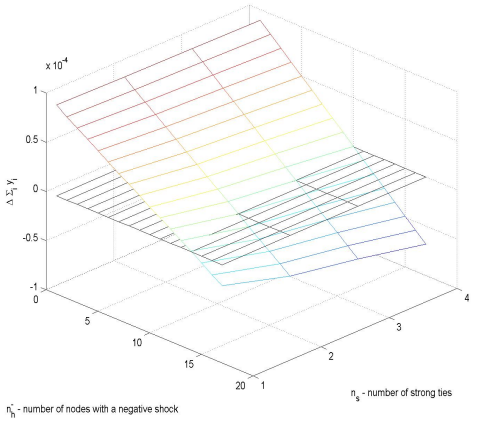
Figure 3.6: Simulation Results
Income Shocks and Strong Tie Effects



(a) Intensive Margin (b) Extensive Margin

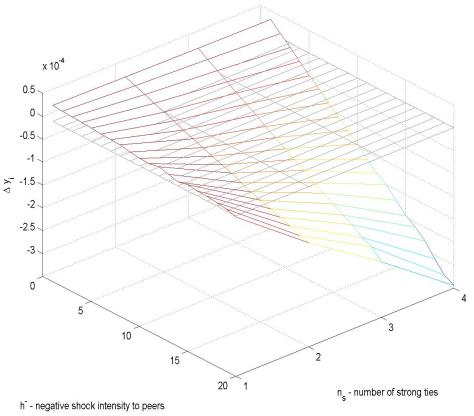
Notes: the surfaces represent $\sum_i \Delta y_i$, which is the variation of the financial activity of agent i, y_i, after the shock. n_s is the number of strong ties of the shocked agents. In Panel (a) shock intensity (h) goes from 1 to 20 income std points, while the number of shocked agents is constant and equal to 13. For each combination of (n_s, h) the income of a random sample of agents which have a n_s strong ties is increased by h. In Panel (b) the shock intensity is constant and equal to 2 income std points, while the number of shocked agents (n_h) goes from 1 to 13. For each combination of (n_s, n_h) the income of n_h agents, which have n_s strong ties, is increased by 2 income std points. Each point of the surfaces is the average of 500 replications, in which agents are randomly sampled. The results remain basically unchanged if we use a different number of shocked agents in panel (a) or a different shock intensity in panel (b).

Figure 3.7: Simulation Results
Heterogeneous Income Shocks and Strong Tie Effects



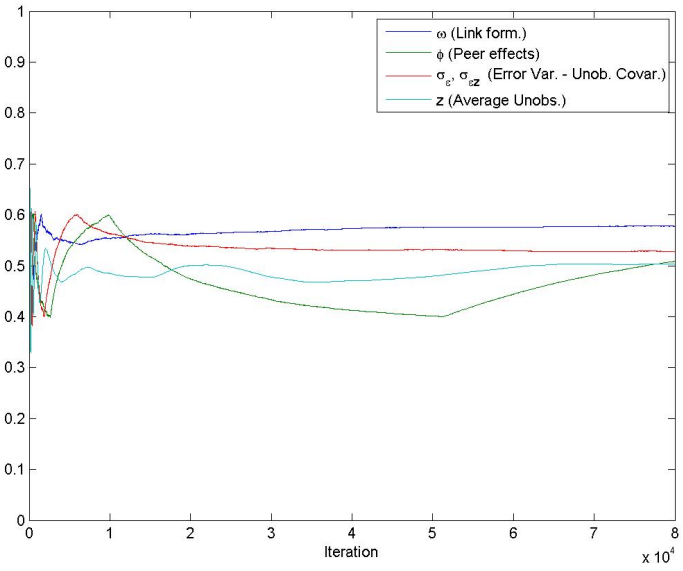
Notes: see Figure 3.6. The income of 13 agents with no strong ties is increased by 2 income std points. The surface represents $\Delta \sum_i y_i$ when the income of n_h^- agents, who have n_s strong ties, is decreased by 2 income std points.

Figure 3.8: Simulation Results
Individual vs Peer Income Shocks



Notes: the surface represents Δy_i, which is the variation of the financial activity of agent i, y_i, after the shock. Each point of the surface is the average of 500 replications in which an agent i is randomly sampled. In each replication, agent i's income is increased by 10 income std points and the income of all of peers of i is decreased by h^- income std points.

Figure B1: Bayesian Estimation Results
Acceptance Rates



Chapter 4

The Allocation of Time in Sleep: a Social Network Model with Sampled Data

Joint work with Xiaodong Liu and Eleonora Patacchini.

Another version of this chapter has been published as *CEPR Discussion Paper No. DP9752*

"Sleep that knits up the ravelled sleave of care, The death of each day's life, sore labour's bath, Balm of hurt minds, great Nature's second course, Chief nourisher in life's feast."

Shakespeare, Macbeth

4.1 Introduction

Nearly a third of a person's life is spent in slumber. In the U.S. those with insomnia spend about $1 billion a year on prescription sleep aids, and another $1 billion on over-the-counter sleep medications Yaniv (2004). The economic costs, both direct (expenditure within the health system) and indirect (absenteeism,

low productivity, and work-related injuries) of sleep disorders in the U.S. in 2004 was estimated to be $109 billion Hillman et al. (2006).

Yet, sleeping behaviour has received relatively little attention in economics. While sleep is primarily a function of the body's internal biological clock (circadian rhythm), individual choice also plays an important role in determining the timing and duration of sleep. Biddle and Hamermesh (1990) posit a simple economic model that accounts for the endogenous nature of sleep choice, but empirical work on the subject has been very limited.

In particular, there is virtually no evidence on the importance of social interactions in shaping sleeping behaviour. In many circumstances, the decision of agents to exert effort in some activity cannot adequately be explained by their personal characteristics and the intrinsic utility derived from the activity. Rather, its rationale may be found in how peers and others value this activity. There is indeed strong evidence that the behaviour of individual agents is affected by that of their peers.[1] The individual utility when allocating time in work or leisure may depend on the same choice made by peers. As a consequence, social interactions might be important for understanding the duration of sleep, which is the residual activity.[2]

In this paper, we exploit the unique information contained in the National Longitudinal Survey of Adolescent Health (AddHealth) to provide evidence on sleeping patterns among adolescents in the U.S. Sleeping behaviour during teenage years is of particular interest because of its effect on human capital formation. Research suggests that lack of sleep reduces attendance, increases tardiness, and lowers grades of adolescent students Eide and Showalter (2012). Furthermore, lack of sleep in youth is correlated with health and behavioral problems such as

[1]The integration of models of social interactions within economic theory is an active and interesting area of research. See the recent *Handbook of Social Economics*, Benhabib et al. (2011).

[2]Biddle and Hamermesh (1990) study the demand for sleep in this perspective without social incentives.

moodiness, depression, difficulty controlling behaviour, and increased frustration - all of which make learning in school difficult (National Sleep Foundation; Mitru (2002)). Sleep also affects productivity on the job, which in some cases represents a public safety concern.

The AddHealth data contain unique information on friendship relationships among a representative sample of students from U.S. high school teenagers together with basic information on individual, family, neighborhood and school characteristics (in-school survey). The survey design also includes a questionnaire administered to a random sample of those students collecting information on more sensitive topics (health issues, crime, drug, sexual behaviour, etc.), including time and duration of sleep on week days during the school year (in-home survey). The use of this additional information, however, comes at a cost. The in-home sampling scheme may result in missing observations on the behaviour of friends who were not sampled, and induce measurement error to the endogenous peer effect variable given by the average behaviour of friends. As a result, the existing estimation methods for network models of social interactions (see, e.g. Bramoullé et al.; 2009; Lee et al.; 2010) are not generally valid.[3]

Recently, social network studies have drawn a great deal of attention. Network models are widely used to represent relational information among interacting units and the implications of these relations. Most inference for social network models assumes that the all possible links are observed and that all the relevant information is available. This is clearly not true in practice, as much network data is collected though sample surveys. In a recent paper, Sojourner (2013) considers a linear-in-means social interaction model with missing observations on covariates. He shows that random assignment of agents to peer groups can help to overcome the missing data problem. On the other hand, Chandrasekhar and Lewis (2011b) consider the estimation of network models with sampled observations on network

[3]This issue is typically neglected in most empirical papers using the information on friends together with the in-home survey in the AddHealth data set.

links. They propose a set of analytical corrections for commonly used network statistics and a two-step estimation procedure using graphical reconstruction. Our case is different. We observe all the network links and the covariates for all nodes, but we have sampled observations on the dependant variables.

The social network model considered in this paper has the specification of a spatial autoregressive (SAR) model with group-specific fixed effect. Kelejian and Prucha (2010) consider the estimation of the SAR model with missing observations on the dependent variable and covariates. They suggest two-stage least squares (2SLS) estimators that are based on a subset of the sample so that the dependent variable and covariates are observed, and the spatial lags are either completely observed or partially observed with an asymptotically negligible measurement error. Our set up is similar to the one proposed by Wang and Lee (2013a). Wang and Lee (2013a) consider the estimation of the SAR model with missing observations on the dependent variable for cross-sectional data Wang and Lee (2013a) and for random effect panel data Wang and Lee (2013b). They propose the generalized method of moments (GMM) estimator, the nonlinear least squares (NLS) estimator, and the 2SLS estimator with imputation. They show that the three estimators are consistent and robust against unknown heteroskedasticity. In this paper, we extend the NLS estimator in Wang and Lee (2013a) to estimate social network models with network fixed effects and sampled observations on the dependent variable. While the asymptotic properties of the proposed estimator is similar to Wang and Lee (2013a) (see Section 4.3.2), this is the first paper applying this approach to real data and estimating network models with sampling.

Our results show that the conventional 2SLS is inconsistent without accounting for sampling. In our case, 2SLS fails to detect the presence of peer effects. When sampling is taken into account, we instead find that the sleeping behaviour of the friends is important in shaping own sleeping behaviour, besides the impact

of individual and friends characteristics. We use the approach recently proposed by Goldsmith-Pinkham and Imbens (2013) to investigate testable implications of network endogeneity, finding no sign of troubling individual level unobservables that may invalidate our results. Our results are also robust when using an unique information on siblings to eliminate possible unobserved family factors.

In summary, we make the following contributions to the existing literature:

i) we evaluate the bias of the traditional 2SLS when estimating a social network model with a sampled dependent variable in small samples using a Monte Carlo experiment;

ii) we develop a correction using a NLS method with network fixed effects, which is easy to implement in applied work;

iii) we provide the first empirical application of the method to a unique dataset of friendship networks finding that young adults respond to the sleeping behavior of their peer group, holding constant other observables. This effect suggests a group approach to solving behavioral problems associated with sleep deprivation;

iv) we propose a methodology to test the presence of implications of network endogeneity;

v) we use data on siblings and their friends to account for unobserved family effects.

The paper is organized as follows. We start our analysis by describing our data in Section 2. Section 3 presents the network model, together with the identification and estimation strategy. We discuss our estimation results in Section 4, whereas Section 5 contains some robustness checks. Section 6 concludes.

4.2 Data and Descriptive Evidence

Our data source is the AddHealth data that has been designed to study the impact of the social environment (i.e. friends, family, neighborhood and school)

on adolescents' behaviour in the United States by collecting data on students in grades 7-12 from a nationally representative sample of roughly 130 private and public schools in years 1994-95. Every student attending the sampled schools on the interview day is asked to compile a questionnaire (in-school survey) containing questions on respondents' demographic and behavioral characteristics, education, family background and friendship. Most notably, students were asked to identify their best friends from a school roster - up to five males and five females. The limit in the number of nominations, however, is not binding (not even by gender),[4] and in the large majority of cases (more than 90%) the nominated best friends are in the same school. Hence, it is possible to reconstruct the entire geometry of the friendship networks within each school. In addition, by matching the identification numbers of the friendship nominations to respondents' identification numbers, one can obtain information on the characteristics of nominated friends. This sample contains information on roughly 90,000 students. These features make these data almost unique. It is extremely rare to have information on the universe of network contacts (here school friends), together with their detailed characteristics.[5] The survey design also includes a longer questionnaire (in-home survey) containing questions related to more sensitive individual and household information which is administered to a subset of adolescents. We use the *core sample* of in-home survey which provides information on a random and self-weighting subset of adolescents, about 12,000 individuals.[6] The in-home questionnaire contains detailed information about the timing and duration of sleep. The questions has been slightly reformulated over time to measure sleeping pat-

[4]Less than 1 percent of the students in our sample show a list of ten best friends, less than 3 percent a list of five males and roughly 4 percent name five females. On average, they declare to have 4.35 friends with a small dispersion around this mean value (standard deviation equal to 1.41).

[5]The information on social network contacts collected in other existing surveys is about "ego-networks", i.e. the respondent is asked to name few personal contacts and provides (self-reported) information about an extremely limited number of their charactersitics.

[6]The *core sample* contains roughly the 60% of the individuals interviewed in the in-home survey (which are about 20,000 individuals). The difference is due to the fact that in the in-home sampling design some types of individuals are oversampled.

terns more precisely. Indeed, the (in-home survey) students are interviewed again one year later, in 1995–96 (wave II).[7] We derive the information on sleeping patterns by using the wave II question: "During the school year, what time do you usually go to bed on week nights?".[8,9]

Figure 1 plots the empirical distribution. The graph shows a notable dispersion around the mean "bed time" value (mean equal to 10:37pm and standard deviation equal to 58.7 minutes). About 50% of the students go to bed between 10pm and 11.30pm.

The implicit assumtion is that very school starts at the same time, so that by looking at the time teenagers go to sleep we can recover their sleeping duration. This is not accurate since schools may start at different times, even within the same school for different grades to minimize cost of busing. However, Edwards (2012) documents that the nationwide dispersion of school start time is low: the 25 – 75 percentiles interval is 7:55 – 8:30 (35 minutes). The much higher dispersion in "bed-time" in our sample cannot be explained only by the school start time. Nevertheless, to address this point we include network fixed effects. Because friendship networks in our sample are within school and grade, by conditioning on network fixed effects we compare people with the same school start time.

Figure 2 shows the distribution of students by GPA distinguishing between students with different sleeping patterns. It appears that students with sleep deficit (red curve) show a statistically significant lower performance at school.[10]

[7]Those subject are also interviewed again in 2001-02 (wave III), and again in 2007-08 (wave IV). For the purposes of this paper, we do not use this longitudinal information. The friendship nominations are only collected when the students were at school (i.e. in waves I and II).

[8]The questions formulated in wave I do not differentiate between the school period and summer time. The same issue also applies to the other question on sleeping behaviour in Wave II: "How many hours of sleep do you usually get?". Finally, a third question is available: "Do you usually get enough sleep?", which measures a subjective perception, thus increasing measurement errors. In addition, the answers to both questions are not continuos variables, as requested by the NLS estimation.

[9]We rescaled each hour in 100 units, so for instance half an hour is transformed to a distance of 50. We dropped individuals declaring going to sleep before 5pm and after 6am.

[10]The rejection of the null hypothesis in a Kolmogorov-Smirnov test confirms the difference between these two distributions.

In other words, a student that goes to bed earlier is more likely to have a higher GPA.

Table 1 and Figure 3 collects some further evidence on the relationship between sleeping patterns and other relevant characteristics. We run a principal component analysis (PCA)[11] on body mass index (BMI), GPA, general health, use of alcohol and cigarette smoking. The first principal component explains over one third of the total inertia. Table 1 shows that this variation is associated to differences between two clusters of students, one with high body mass index, poor school performance, poor general health, drinking alcohol and smoking cigarettes (type A students), and the other with the opposite profile (type B students). In other words, splitting the population between type A and type B individuals maximizes the between-group variation and minimizes the within-group variation. Figure 3 shows that type A students tend to sleep for fewer hours than type B students. This is in line with an (expected) relationship between sleeping behaviour and individual socio-economic profile Eide and Showalter (2012).

4.3 Regression Analysis

Our aim is to assess the actual empirical relationship between the individual sleeping behaviour and the sleeping behaviour of the peers using the unique information provided by the AddHealth data. This exercise requires facing the traditional challenges in identifying endogenous social interaction effects, while also overcoming a further (and so far neglected) issue stemming from the sampling design of the AddHealth survey. We present the network model in Section 3.1, whereas the estimation of network models with sampling on the dependant variable is considered in detail in Section 3.2.

[11] PCA uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables (called principal components). This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for the largest portion of variability in the data).

4.3.1 The network model

Consider a population of n individuals partitioned into $\bar{r}$ networks. For the n_r individuals in the rth network, their connections with each other are represented by an $n_r \times n_r$ adjacency matrix $G_r^* = [g_{ij,r}^*]$ where $g_{ij,r}^* = 1$ if individuals i and j are friends and $g_{ij,r}^* = 0$ otherwise.[12] Let $G_r = [g_{ij,r}]$ be the row-normalized G_r^* such that $g_{ij,r} = g_{ij,r}^* / \sum_{k=1}^{n_r} g_{ik,r}^*$.

Given the network adjacency matrix G_r, we assume $y_{i,r}$, the sleeping behaviour of individual i in network r, is given by the following network model

$$y_{i,r} = \phi \sum\nolimits_{j=1}^{n_r} g_{ij,r} y_{j,r} + \sum\nolimits_{k=1}^{p} x_{ik,r}\beta_k + \sum\nolimits_{k=1}^{p} (\sum\nolimits_{j=1}^{n_r} g_{ij,r} x_{jk,r}\gamma_k) + \eta_r + \epsilon_{i,r}. \quad (4.1)$$

In this model, $\sum_{j=1}^{n_r} g_{ij,r} y_{j,r}$ is the average sleeping behaviour of i's direct friends with its coefficient ϕ representing *the endogenous effect*, wherein an individual's choice/outcome may depend on those of his/her friends about the same activity. $x_{ik,r}$, for $k = 1, \cdots, p$, are exogenous control variables. For $k = 1, \cdots, p$, $\sum_{j=1}^{n_r} g_{ij,r} x_{jk,r}$ is the average value of the k-th control variable taking over i's direct friends with its coefficient γ_k representing *the contextual effect*, wherein an individual's choice/outcome may depend on the exogenous characteristics of his/her friends. η_r is a network-specific parameter representing *the correlated effect*, wherein individuals in the same group tend to behave similarly because they face a common environment. $\epsilon_{i,r}$ is an i.i.d. error term with zero mean and finite variance σ^2.

Let $x_{i,r} = (x_{i1,r}, \cdots, x_{ip,r})'$, $\beta = (\beta_1, \cdots, \beta_p)'$ and $\gamma = (\gamma_1, \cdots, \gamma_p)'$. In matrix form, (4.1) can be rewritten as

$$Y_r = \phi G_r Y_r + X_r\beta + G_r X_r \gamma + \eta_r l_{n_r} + \epsilon_r, \quad (4.2)$$

[12] For ease of presentation, we focus on the case where the connections are undirected and no agent is isolated so that G_r^* is symmetric and $\sum_{j=1}^{n} g_{ij,r}^* \neq 0$ for all i. The result of the paper holds for a directed network with an asymmetric G_r^*.

where $Y_r = (y_{1,r}, \cdots, y_{n_r,r})'$, $X_r = (x_{1,r}, \cdots, x_{n_r,r})'$, $\epsilon_r = (\epsilon_{1,r}, \cdots, \epsilon_{n_r,r})'$, and l_{n_r} is an $n_r \times 1$ vector of ones.

Let $\text{diag}\{A_j\}_{j=1}^m$ denote a generalized diagonal block matrix with the diagonal blocks being A_j's, where A_j may or may not be a square matrix. Then, for all $\bar{r}$ networks, we can stack the data such that (4.3) becomes

$$Y = \phi GY + X\beta + GX\gamma + L\eta + \epsilon, \tag{4.3}$$

where $Y = (Y_1', \cdots, Y_{\bar{r}}')'$, $G = \text{diag}\{G_r\}_{r=1}^{\bar{r}}$, $X = (X_1', \cdots, X_{\bar{r}}')'$, $L = \text{diag}\{l_{n_r}\}_{r=1}^{\bar{r}}$, $\eta = (\eta_1, \cdots, \eta_{\bar{r}})'$, and $\epsilon = (\epsilon_1', \cdots, \epsilon_{\bar{r}}')'$.

The identification and estimation of endogenous, contextual, and correlated effects have been the main interests of social network models. The conventional identification and estimation strategy in the literature (see, e.g. Lee; 2007; Bramoullé et al.; 2009; Lee et al.; 2010) relies on the assumption that $\text{E}(\epsilon_r | G_r, X_r, \eta_r) = 0$.[13] Based on this assumption, Bramoullé et al. (2009) show that if intransitivities exist in networks so that I_n, G, G^2, G^3, are linearly independent, then model (4.2) is identified. For estimation, we first eliminate the incidental parameters η using a within-transformation projector $J = \text{diag}\{J_r\}_{r=1}^{\bar{r}}$, where $J_r = I_{n_r} - \frac{1}{n_r} l_{n_r} l_{n_r}'$. As $JL = 0$, premultiplying (4.3) by J, we have

$$JY = \phi JGY + JX\beta + JGX\gamma + J\epsilon.$$

Let $Z = (GY, X, GX)$ and $\theta = (\phi, \beta', \gamma')'$. For the instrumental variable (IV) matrix $Q = (X, GX, G^2X)$, the two-stage least squares estimator is given by

$$\hat{\theta}_{2sls} = (\hat{Z}' JZ)^{-1} \hat{Z}' JY, \tag{4.4}$$

where $J\hat{Z} = JQ(Q'JQ)^{-1}Q'JZ$ is the predicted JZ from the first-stage regres-

[13] We will investigate the validity of this assumption for this empirical study in Section 4.5.

sion.

In the following section, we focus on the sampling issue of the network model that has been largely ignored by the literature.

4.3.2 Estimation of peer effects with sampling

In our and many other studies, the analysis of the network model (4.1) has been made possible by the use of a unique database on friendship networks from the AddHealth data.[14] As we explain in Section 2, students are asked to identify their best friends from the school roster in the in-school survey. Thus, we can observe all friendship links in the networks. However, as some more sensitive individual information - (i.e. sleeping behaviour) - is in the in-home survey, we only have this information for the sampled students.[15]

Without loss of generality, suppose the first m_r ($m_r > 1$) individuals in network r are sampled. Suppose we can observe network connections $G_r = [g_{ij,r}]$ and controls $x_{i,r}$ for all individuals in network r, but we can only observe $y_{i,r}$'s of sampled individuals. For the sampled individuals, $i = 1, \cdots, m_r$, (4.1) becomes

$$y_{i,r} = \phi \sum\nolimits_{j=1}^{m_r} g_{ij,r} y_{j,r} + x'_{i,r}\beta + \sum\nolimits_{j=1}^{n_r} g_{ij,r} x'_{j,r}\gamma + \eta_r + \epsilon^*_{i,r}. \tag{4.5}$$

By comparing (4.1) and (4.5), we have $\epsilon^*_{i,r} = \phi \sum_{j=m_r+1}^{n_r} g_{ij,r} y_{j,r} + \epsilon_{i,r}$. Therefore, the error term of model (4.5) contains two types of errors - the error due to unobserved individual heterogeneity $\epsilon_{i,r}$ and the measurement error due to the sampling design $\phi \sum_{j=m_r+1}^{n_r} g_{ij,r} y_{j,r}$. The measurement error could be correlated with the control variables and, as a result, the 2SLS given by (4.4) may not be consistent.

[14] See, e.g. Lin (2010a), Patacchini and Zenou (2008) and the references herein.

[15] The use of the *core sample* is crucial because otherwise the sampled students are not random.

To further illustrate this point, we rewrite (4.5) in matrix form. Let

$$G_r = \begin{bmatrix} G_r^S \\ G_r^N \end{bmatrix} = \begin{bmatrix} G_r^{SS} & G_r^{SN} \\ G_r^{NS} & G_r^{NN} \end{bmatrix},$$

where G_r^S is an $m_r \times n_r$ matrix of the first m_r rows of G_r and G_r^{SS} is an $m_r \times m_r$ matrix of the first m_r columns of G_r^S. Then, for the sampled individuals, we have

$$Y_r^S = \phi G_r^{SS} Y_r^S + X_r^S \beta + G_r^S X_r \gamma + \eta_r l_{m_r} + \epsilon_r^*, \tag{4.6}$$

where $Y_r^S = (y_{1,r}, \cdots, y_{m_r,r})'$ denotes the $m_r \times 1$ vector of observations on the dependent variable of the sampled individuals, $X_r^S = (x_{1,r}, \cdots, x_{m_r,r})'$ denotes the $m_r \times p$ matrix of observations on the control variables of the sampled individuals, and $\epsilon_r^* = \epsilon_r^S + \phi G_r^{SN} Y_r^N$ with $\epsilon_r^S = (\epsilon_{1,r}, \cdots, \epsilon_{m_r,r})'$ and $Y_r^N = (y_{m_r+1,r}, \cdots, y_{n_r,r})'$. As $\mathrm{E}(\epsilon_r | G_r, X_r, \eta_r) = 0$, we have

$$\mathrm{E}(\epsilon_r^* | G_r, X_r, \eta_r) = \mathrm{E}(\epsilon_r^S + \phi G_r^{SN} Y_r^N | G_r, X_r, \eta_r) = \phi G_r^{SN} \mathrm{E}(Y_r^N | G_r, X_r, \eta_r).$$

To obtain $\mathrm{E}(Y_r^N | G_r, X_r, \eta_r)$, we need to inspect the reduced form equation of the model. If $(I_{n_r} - \phi G_r)$ is nonsingular, the reduced form equation of (4.2) is given by

$$Y_r = (I_{n_r} - \phi G_r)^{-1}(X_r \beta + G_r X_r \gamma) + \frac{\eta_r}{1-\phi} l_{n_r} + (I_{n_r} - \phi G_r)^{-1} \epsilon_r. \tag{4.7}$$

Let $D_r^N = [0_{(n_r - m_r) \times m_r}, I_{n_r - m_r}]$ denote an $(n_r - m_r) \times n_r$ matrix of the last $(n_r - m_r)$ rows of an identity matrix. Then, it follows from (4.7) that

$$\mathrm{E}(Y_r^N | G_r, X_r, \eta_r) = D_r^N \mathrm{E}(Y_r | G_r, X_r, \eta_r) = D_r^N (I_{n_r} - \phi G_r)^{-1}(X_r \beta + G_r X_r \gamma) + \frac{\eta_r}{1-\phi} l_{n_r - m_r}.$$

Therefore,

$$\mathrm{E}(\epsilon_r^*|G_r, X_r, \eta_r) = \phi G_r^{SN}\mathrm{E}(Y_r^N|G_r, X_r, \eta_r) = \phi G_r^{SN} D_r^N (I_{n_r} - \phi G_r)^{-1}(X_r\beta + G_r X_r\gamma) + \frac{\phi\eta_r}{1-\phi} G_r^{SN} l_{n_r -}$$

As $\mathrm{E}(\epsilon_r^*|G_r, X_r, \eta_r)$ is not zero in general, the 2SLS estimator given by (4.4) may not be consistent for (4.6).

To avoid the measurement error due to sampling, we consider the NLS approach suggested by Wang and Lee (2013a) based on the reduced form equation (4.7). Let $D_r^S = [I_{m_r}, 0_{m_r \times (n_r - m_r)}]$ be an $m_r \times n_r$ matrix of the first m_r rows of an identity matrix. Then,

$$Y_r^S = D_r^S Y_r = D_r^S (I_{n_r} - \phi G_r)^{-1}(X_r\beta + G_r X_r\gamma) + \frac{\eta_r}{1-\phi} l_{m_r} + u_r, \tag{4.8}$$

where $u_r = D_r^S (I_{n_r} - \phi G_r)^{-1}\epsilon_r$. As $\mathrm{E}(u_r|G_r, X_r, \eta_r) = 0$, a regression estimator based on (4.8) would be consistent.

First, to eliminate the incidental parameters η_r, we apply a within transformation using the projector $J_r^S = I_{m_r} - \frac{1}{m_r} l_{m_r} l'_{m_r}$ so that (4.8) becomes

$$J_r^S Y_r^S = J_r^S h_r(\theta) + J_r^S u_r,$$

where $h_r(\theta) = D_r^S (I_{n_r} - \phi G_r)^{-1}(X_r\beta + G_r X_r\gamma)$ with $\theta = (\phi, \beta', \gamma')'$. The NLS estimator of θ is given by

$$\hat{\theta}_{nls} = \arg\min_{\theta} \sum\nolimits_{r=1}^{\bar{r}} [Y_r^S - h_r(\theta)]' J_r^S [Y_r^S - h_r(\theta)]. \tag{4.9}$$

Let $J^S = \mathrm{diag}\{J_r^S\}_{r=1}^{\bar{r}}$ and $D^S = \mathrm{diag}\{D_r^S\}_{r=1}^{\bar{r}}$. Following a similar argument in Wang and Lee (2013a), the NLS estimator $\hat{\theta}_{nls}$ is consistent with an asymptotic distribution

$$\sqrt{n}(\hat{\theta}_{nls} - \theta) \xrightarrow{d} N(0, \Sigma_{nls}),$$

where $\Sigma_{nls} = \lim_{n\to\infty} n(C'B'BC)^{-1}C'B'\Omega BC(C'B'BC)^{-1}$, with $B = J^S D^S(I - \phi G)^{-1}$, $C = [G(I - \phi G)^{-1}(X\beta + GX\gamma), X, GX]$ and $\Omega = \sigma^2 BB'$.[16]

4.3.3 A simulation experiment

We conduct a Monte Carlo simulation in which we compare the 2SLS estimator which is commonly used for the estimation of peer effects and the NLS estimator given in (4.9). The setup of our simulations is as follows. The population numerosity is 500 nodes and the number of separated networks is 50, resulting in subnetworks of 10 nodes. Each node is allowed to have three connections as a maximum and zero as a minimum with a uniform distribution within the subnetwork to which it belongs. Links are formed randomly. We consider sampling rates of 40 percent, 60 percent, 80 percent, 100 percent . For each rate and for each estimator, we estimate 5,000 times model (4.1) using one variable x. The control variable x and the network fixed effect η are randomly generated by a normal distribution $N(0, 1)$. The innovation ϵ is generated by a normal distribution $N(0, \sigma^2)$. We set $\lambda = 0.3$, $\beta = 1.0$, $\gamma = 1.0$, and $\sigma^2 = 2$ in the data generating process.[17] Table 2 reports the results of our Monte Carlo study. The NLS estimates roughly coincide with the true parameter values. The 2SLS estimates are downwards biased, with the magnitude of the bias increasing as the sampling rate decreases. The NLS and 2SLS have similar performance when all individuals are sampled (i.e. the sampling rate is zero). We have also repeated our simulations when varying the maximum number of connections (i.e. the network density) and using various distributions (other from uniform). The results are stable across the different specifications.[18]

[16] As in Wang and Lee (2013a), we assume the number of sampled individuals is proportional to n so that the convergence rate of the estimator can be written in terms of n.

[17] Conclusions of our simulation study are not sensitive to the parameters values. For the sake of brevity we do not show the output of all simulations.

[18] We do not report these further results for brevity. They remain available upon request.

4.4 Estimation Results

Having in mind the simulation results, we move to the empirics and follow the same comparative approach among different methods.

Our main estimation results are reported in Table 3. The dependent variable is the time students go to bed. During the school days, this variable captures the time allocated to sleep - the later a student goes to bed, the lower is her/his sleep duration. The different columns show the results with an increasing set of controls. In the first specification, we include individual demographic characteristics, family background characteristics, contextual effects (the average of peers' characteristics) and network fixed effects. We introduce scores in mathematics and history/social science in the second specification, and finally we include a risky behaviour factor in the third specification.[19] The results can be summarized as follows.

First, with the exception of peer effects, point estimates and standard errors are stable across specifications and estimators. The results are in line with the expectations. Biddle and Hamermesh (1990) model the demand for sleep as a function of wage and leisure. In their model, the higher the value of an additional worked hour (i.e. the higher the wage), the lower is the time allocated to sleep. Although we deal with students rather than workers, the general mechanisms still apply. If one interprets the return of school performance as wage, then we expect a negative correlation between student grade and sleep duration because incentives to spend hours in studying increase over the school years. Similarly, if time spent in risky behaviour is seen as leisure time, then an increase in risky activities should negatively impact the amount of time allocated to sleep.

Second, the peer effect estimated coefficient is significantly different from zero for all specifications when estimated using the NLS estimator, while it is never

[19]The Risky Behavior Factor is the score of a factor analysis run on use of alchool, cigarette smoking and general health. The results are robust to alternativfe sets of controls.

significantly different from zero when using 2SLS. In addition, our estimator shows both point estimates and standard errors which are stable across specifications. In terms of magnitude, in the average group of four people, an additional hour of sleep of each of the friends translates to about 45 minutes in the individual sleeping duration.

Note that this empirical evidence is in line with the simulation results, since the downwards bias here leads the 2SLS to suggest that no peer effect is at work, unlike with NLS.

In order to better understand the magnitude of the effects, we provide an evaluation of the individual response to peers' sleeping variations in terms of child's school performance. Our results reveal that in a group of friends, an additional hour of sleep of each friend translates into about 45 minutes in the individual sleeping duration. A unit increase in the individual sleeping duration is associated with a 0.02% increase in GDP,[20] which translates into about 1.5% for a 45 minute increase. Therefore, if a friend sleeps one more hour, then the individual school performance would increase by roughly 1,5%. This is a small effect, as it is probably expected.

4.5 Robustness Checks

4.5.1 Endogenous network formation

An important feature of our identification strategy is the use of network fixed effects. In most cases individuals sort into groups non-randomly. For example, kids whose parents are low educated or worse than average in unmeasured ways would be more likely to sort with low human capital peers. If the variables that drive this process of selection are not fully observable, potential correlations between (unobserved) group-specific factors and the target regressors are major

[20] We run a simple regression - (log) gdp on sleep duration and controls .

sources of bias. It is thus difficult to disentangle the endogenous peer effects from the correlated effects, i.e. from effects arising from the fact that individuals in the same group tend to behave similarly because they face a common environment. Network fixed effect are a remedy for the selection bias that originates from the possible sorting of individuals with similar unobserved characteristics into a network. The underlying assumption is that such unobserved characteristics are common to the individuals within each network. This is reasonable in our case study where the networks are quite small (see Section 2). However, if there are student-level unobservables that drive both network formation and outcome choice, then G_r is endogenous and this strategy fails.

We provide here some evidence that helps to reveal whether in our model G_r violates the exogeneity condition. The intuition behind the test is to evaluate whether differences in unobserved factors that drive outcome decisions help also to explain whether those individuals are friends. Evidence of correlation would be suggestive that unobserved individual characteristics shape network formation as well as eduaction achievement which could point to a potential violation of exogeneity.

Let us consider a network formation model based on homophily behaviors where the variables that explain $g_{ij,r}$ are distances in terms of observed and unobserved characteristics between students i and j:[21]

$$g^*_{ij,r} = \alpha + \sum_{m=1}^{M} \delta_m x^m_{ij,r} + \theta v_{ij,r} + \eta_r + u_{ij,r}, \tag{4.10}$$

where $g^*_{ij,r}$ is the latent variable, $x^m_{ij,r}$ is 1 if x^m is a discrete variable and i and j have the same value (0 otherwise) while it is equal to $1/|x^m_{i,r} - x^m_{j,r}|$ if x^m is continuous, $v_{ij,r}$ is equal to $1/|v_{i,r} - v_{j,r}|$, $v_{i,r}$ and $v_{j,r}$ are the residuals from the outcome equation (4.8) for individual i and j respectively. θ is the parameter of

[21] See Jackson (2008) for a discussion on the theorical nature of such model and Currarini et al. (2009) for an empirical application.

interest.

Estimates of expression (4.10) are presented in in Table 4.[22] The results show no sign of correlation between differences in unobserved individual characteristics and link formation. It should also be noted that because there are several thousands individual-pair observations in a given regression, the power to detect small departures from zero is quite high.

As a result, conditional on the (unusually) large set of individual characteristics provided by the AddHealth, peer characteristics and network effects, we find no evidence of network endogeneity.[23]

4.5.2 Siblings

Let us conclude our analysis with a further robustness check.

The restricted-use version of the AddHealth dataset contains sibling pairs data. For each respondent, we know who is the sibling, her/his characteristics, the nominated friends and her/her friends' characteristics. We exploit this unique source of information to test whether peer effects are still significantly different from zero if we introduce sibling fixed effects. If our peer effect estimate is simply picking up unobserved individual characteristics, then we should find no effect when washing away the influence of factors that are common for siblings who grew up in the same family and consequently have been educated by the same persons, lived in the same neighborhood and more generally faced a wide number of common shocks.

Almost all our sample of siblings (about 97%) are in the same social network, i.e. are indirectly connected through a chain of friends. However, they have different direct friends. So this is the source of variation which is exploited in our

[22] Both a linear probability and a logit model are estimated.

[23] Goldsmith-Pinkham and Imbens (2013) suggest the use of a similar diagnostic procedure to investigate a possible endogenous formation of networks prior to the use of Bayesian estimation techniques. Indeed, signals of individual-level correlated unobservables would motivate the use of parametric modeling assumptions and Bayesian inferential methods to integrate a network formation with the study of behaviour over the formed networks.

sibling fixed effect strategy.

Table 5 shows the estimation results. The coefficient estimates are reduced in magnitude and the parameters are less precisely estimated due to the reduced sample size. However, the substance of the results remain unchanged: the peer effect estimate remains significantly different from zero when using the NLS estimator in all specifications.

4.6 Conclusions

There is remarkably little evidence on the determinant of individual differences in sleep duration. By implementing sound econometric techniques, our study is able to provide novel evidence in this respect. We have two contributions to the literature. One, we extend the NLS estimator in Wang and Lee (2013a) to estimate social network models with sampled observations on the dependent variable. Two, we analyze peer effects in sleeping behaviour using a representative sample of U.S. teenagers, finding not-negligible endogenous effects. That is, besides the impact of individual and friend characteristics, we show that the sleeping behaviour of the friends is important in shaping own sleeping behaviour. Unique information on siblings and their friends allows us to check the robustness of our results to unobserved family factors.

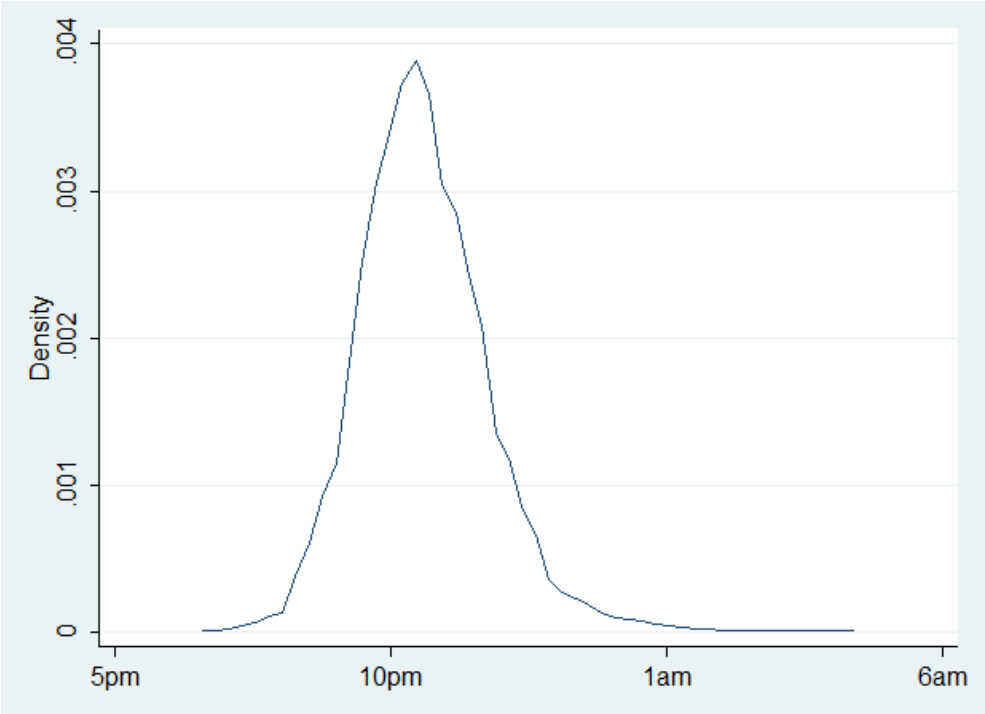
Adolescent sleep patterns deserve particular attention because of their potential to affect school performance. Side effects associated with sleep deprivation - inattention, irritability, hyperactivity, and impulse control problems - are likely to show up in school. It is important for educators to screen for sleep problems when concerns exist about a student's attention or behavior problems. Our analysis suggests that an effective intervention should not only be measured by the possible sleep disorder reduction it implies but also by the group interactions it engenders.

Table 4.1: Student Characteristics - PCA results -

Variable	Correlation with the first PC
Body mass Index	0.19
GPA	-0.32
General Health	-0.39
Use of Alchool	0.56
Cigarette smoking	0.57

Notes. The first PC explain 32% of the total variance. Body mass index is the ratio between weight (in kilos) and the height squared (in meters). GPA is the composite score of a factor analysis run on Mathematics score, English score, History/Social Science score and Science score. General health is derived from the question: "In general how is your health?", coded as 1= excellent, 2= very good, 3= good, 4= fair, 5= poor. The use of alchool is measured using the question: "During the past twelve months, how often did you: drink beer, wine, or liquor?", coded as 0 = never, 1= once or twice, 2= once a month or less, 3= 2 or 3 days a month, 4= once or twice a week, 5= 3 to 5 days a week, 6= nearly every day. Cigarette smoking uses responses to the question: "During the past twelve months, how often did you: smoke cigarettes?", coded as 0 = never, 1= once or twice, 2= once a month or less, 3= 2 or 3 days a month, 4= once or twice a week, 5= 3 to 5 days a week, 6= nearly every day.

Figure 4.1: Kernel Density Estimate of Bed Time



Notes: Kernel = Epanechnikov, bandwidth = 40.429. We report the distribution of student by the time they go to sleep.

Table 4.2: Simulation Results

Sampling rate	Method	Parameter	Point estimation	Standard error	MSE
40%					
		λ	0.293	0.152	0.152
	NLS	β	1.002	0.093	0.093
		γ	1.013	0.179	0.180
		λ	0.255	0.546	0.548
	2SLS	β	0.995	0.120	0.120
		γ	1.007	0.535	0.535
60%					
		λ	0.294	0.117	0.117
	NLS	β	1.001	0.072	0.072
		γ	1.011	0.138	0.139
		λ	0.252	0.216	0.221
	2SLS	β	0.997	0.081	0.081
		γ	1.013	0.222	0.222
80%					
		λ	0.294	0.100	0.101
	NLS	β	1.001	0.062	0.062
		γ	1.009	0.116	0.116
		λ	0.272	0.132	0.135
	2SLS	β	0.999	0.066	0.066
		γ	1.007	0.141	0.141
100%					
		λ	0.295	0.090	0.093
	NLS	β	1.001	0.050	0.054
		γ	1.008	0.100	0.103
		λ	0.300	0.091	0.091
	2SLS	β	1.000	0.054	0.054
		γ	1.002	0.101	0.101

Notes. Number of replications =5000. Sample size = 500. Number of groups = 50. Number of nodes per group = 10. Maximum number of connections for a node = 3. Distribution of nodes' connections: uniform. Model: y = λGy + β x + γGx+ ε . λ= 0.3, β = 1.0, γ = 1.0, σ^2 = 2. MSE=$\sqrt{(\theta-\hat{\theta})^2+\mathrm{var}(\hat{\theta})}; \theta=\lambda,\beta,\gamma.$

Table 4.3: Peer effect Estimation Different method comparison- Increasing set of controls

Variable	NLS			2SLS		
Peer effect	0.723**	0.753**	0.726**	-0.213	-0.435	-0.309
	(0.328)	(0.336)	(0.367)	(0.186)	(0.300)	(0.196)
Female	5.579	4.628	5.217	-1.632	-1.350	-1.684
	(5.470)	(5.520)	(5.566)	(5.235)	(5.247)	(5.152)
Grade	27.000***	24.501***	26.812***	24.173***	23.970***	22.780***
	(5.694)	(5.668)	(5.652)	(2.907)	(2.787)	(2.739)
Black	14.624	17.895	22.084	16.208	19.995	22.707
	(16.188)	(16.121)	(16.029)	(12.509)	(12.595)	(15.402)
Asian	20.565	20.006	25.703	24.249	25.255	26.762
	(16.125)	(15.976)	(15.943)	(15.151)	(15.197)	(14.933)
Mathematics score		7.290**	8.491**		9.417***	10.657***
		(3.331)	(3.377)		(3.376)	(3.338)
History/Social Science score		-10.675***	-9.669***		-10.742***	-8.622***
		(3.434)	(3.490)		(3.417)	(3.358)
Risky Behavior Factor			9.557***			10.373***
			(2.518)			(2.343)
Family Characteristics	yes	yes	yes	yes	yes	yes
Contextual effects	yes	yes	yes	yes	yes	yes
Network fixed effects	yes	yes	yes	yes	yes	yes
1,127 Sampled individuals over 3,700 Individuals in 77 Networks						

Notes: Robust standard errors in parentheses, *** p<0,01, ** p<0,05, * p<0,1. Family characteristics include occupation and education of the parents, household size as measured by the number of people living in the household, and a dummy taking value one if the respondent lives in a household with two parents (both biological and non biological) that are married. Parental education is the schooling level of the (biological or non-biological) parent who is living with the child, distinguishing between "never went to school", "not graduate from high school", "high school graduate", "graduated from college or a university", "professional training beyond a four-year college", coded as 0 to 4. We consider only the education of the father if both parents are in the household. Mather and father occupation dummies include the following categories: manager, professional/technical, officer or sales worker, military or security, farm or fishery, other. "None" is the reference group The Risky Behavior Factor is the score of a factor analysis run on use of alcohol cigarette smoking and general health (see the notes to Table 1 for the definition of these variables).

Table 4.4: Endogeneous network formation
-Testable implications

	OLS	Logit
Variable	g_{ij}	g_{ij}
Residuals	-0.0013	-0.3889
	(0.001)	(0.299)
Female	0.0048	0.1938**
	(0.003)	(0.081)
Grade	0.1076***	2.9879***
	(0.012)	(0.100)
Black	0.0120**	0.4987***
	(0.005)	(0.180)
Asian	0.0023	0.1963
	(0.009)	(0.222)
Mathematics score	0.0058	0.1384
	(0.005)	(0.084)
History/Social Science score	0.0088**	0.2924***
	(0.004)	(0.082)
Risky Behavior Factor	0.0159***	0.4900***
	(0.005)	(0.117)
Family Characteristics	Yes	Yes
Observations	23,926	23,926

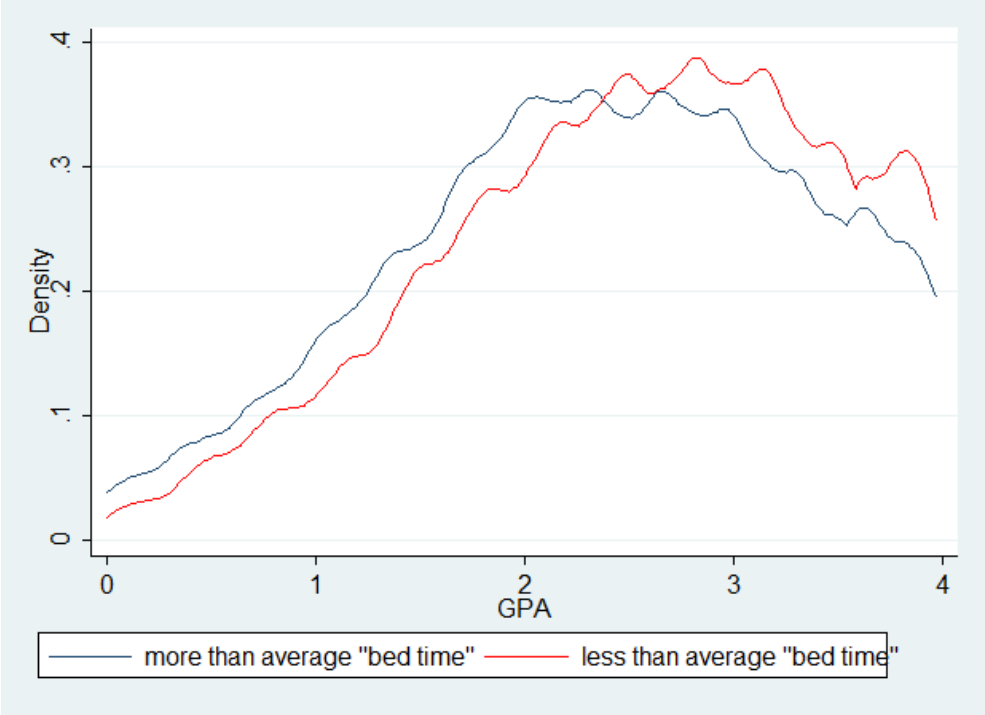
Notes: standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1. Network fixed effects are included.

Table 4.5: Robustness Check. Peer Effect Estimation with Sibling Fixed Effects

Variable	NLS			2SLS		
Peer effect	0,614*	0,547*	0,667*	0,653	0,439	0,128
	(0,319)	(0,298)	(0,395)	(1,063)	(0,895)	(0,943)
Sex	18,243	10,507	-1,433	14,369	12,908	12,321
	(16,608)	(17,126)	(16,901)	(14,351)	(14,361)	(14,272)
Grade	28,115***	26,989***	30,348***	30,817***	28,976***	27,498***
	(6,045)	(8,180)	(7,950)	(5,080)	(4,729)	(4,736)
Black	57,389	29,551	115,329	-10,817	-7,852	-8,849
	(101,692)	(105,599)	(106,494)	(20,428)	(20,016)	(19,927)
Asian	153,274*	90,641	132,325	76,334*	71,505*	70,623*
	(78,011)	(79,331)	(79,610)	(39,991)	(39,423)	(39,131)
Mathematics score		4,332	10,200		13,557*	18,430**
		(9,796)	(10,216)		(8,151)	(8,585)
History/Social Science score		-15,418	-5,331		-5,764	-5,479
		(11,049)	(10,787)		(8,467)	(8,431)
Risky Behavior Factor			12,334			10,526
			(7,743)			(6,825)
Family Characteristics	yes	yes	yes	yes	yes	yes
Contextual effects	yes	yes	yes	yes	yes	yes
Sibling fixed effects	yes	yes	yes	yes	yes	yes
171 Sampled individuals over 3,700 Individuals in 77 Networks						

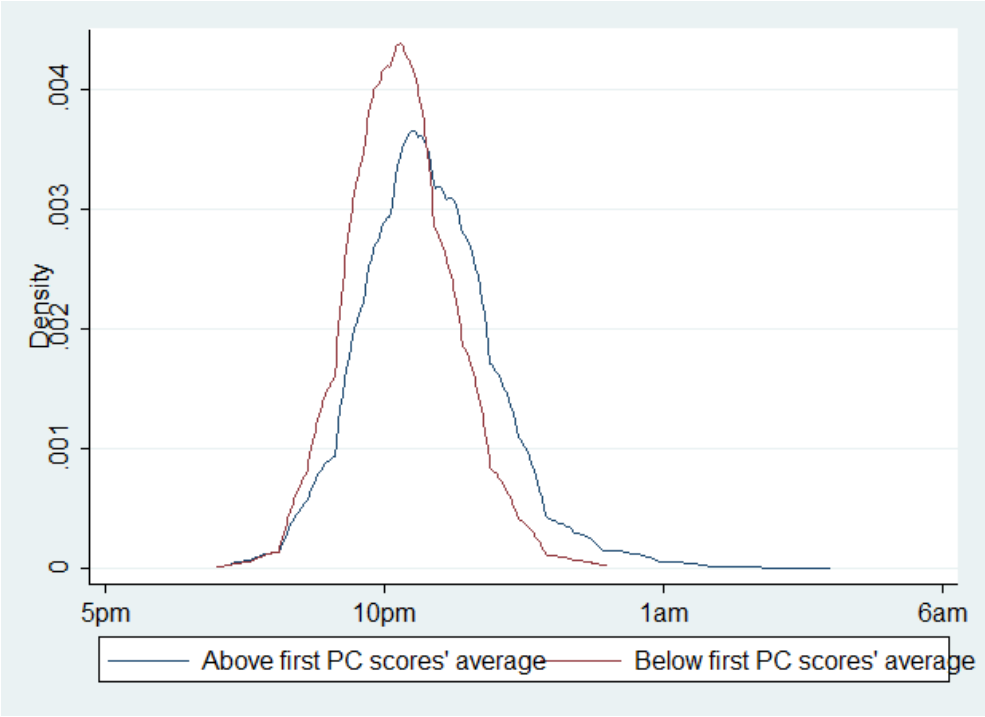
Notes: Robust standard errors in parentheses, *** p<0,01, ** p<0,05, * p<0,1. Family characteristics include occupation and education of the parents, household size as measured by the number of people living in the household, and a dummy taking value one if the respondent lives in a household with two parents (both biological and non biological) that are married. Parental education is the schooling level of the (biological or non-biological) parent who is living with the child, distinguishing between "never went to school", "not graduate from high school", "high school graduate", "graduated from college or a university", "professional training beyond a four-year college", coded as 0 to 4. We consider only the education of the father if both parents are in the household. Mather and father occupation dummies include the following categories: manager, professional/technical, officer or sales worker, military or security, farm or fishery, other. "None" is the reference group The Risky Behavior Factor is the score of a factor analysis run on use of alcohol cigarette smoking and general health (see the notes to Table 1 for the definition of these variables).

Figure 4.2: Bed Timeand School Performance



Notes: Kernel = Epanechnikov, bandwidth = 17.793. We report the distributions of students by school performance as measured by GPA, distinguishing between students that sleep more and less than average. GPA is the composite score of a factor analysis run on Mathematics score, English score, History/Social Science score and Science score.

Figure 4.3: Bed Timeand First PC



Notes: Kernel = Epanechnikov, bandwidth = 40.429. We report the distributions of Type A students (blu line) and Type B students (red line). Type A students have high body mass index, poor school performance, poor general health, drink alcohol and smoke cigarettes, whereas Type B students have the opposite profile.

Bibliography

Abel, A. B. (1990). Asset prices under habit formation and catching up with the joneses, *The American Economic Review* pp. 38–42.

Ahn, I. and Suominen, M. (2001). Word-of-Mouth Communication and Community Enforcement, *International Economic Review* **42**(2): 399–415.

Algan, Y. and Cahuc, P. (2014). Chapter 2 - trust, growth, and well-being: New evidence and policy implications, *in* P. Aghion and S. N. Durlauf (eds), *Handbook of Economic Growth*, Vol. 2 of *Handbook of Economic Growth*, Elsevier, pp. 49 – 120.

Anatolyev, S. (2013). Instrumental variables estimation and inference in the presence of many exogenous regressors, *The Econometrics Journal* **16**(1): 27–72.

Angelucci, M., De Giorgi, G., Rangel, M. and Rasul, I. (2010). Family networks and school enrolment: Evidence from a randomized social experiment, *Journal of Public Economics* **1994**(34): 197–221.

Balmaceda, F. and Escobar, J. (2013). Trust in cohesive communities, *Technical report*, Manuscript.

Baltagi, B. H., Kao, C. and Liu, L. (2012). On the estimation and testing of fixed effects panel data models with weak instruments, *Advances in Econometrics* **30**: 199–235.

Bekker, P. A. (1994). Alternative approximations to the distributions of instrumental variable estimators, *Econometrica* pp. 657–681.

Benhabib, J., Bisin, A. and Jackson, M. O. (eds) (2011). *Handbook of Social Economics*, Vol. 1A, North-Holland.

Benjamini, I. and Peres, Y. (1994). Markov chains indexed by trees, *The annals of probability* **22**: 219–243.

Biddle, J. E. and Hamermesh, D. S. (1990). Sleep and the allocation of time, *Journal of Political Economy* **98**: 922–943.

Bikhchandani, S. and Sharma, S. (2000). Herd behavior in financial markets: A review.

Bramoullé, Y., Djebbari, H. and Fortin, B. (2009). Identification of peer effects through social networks, *Journal of econometrics* **150**(1): 41–55.

Bursztyn, L., Ederer, F., Ferman, B. and Yuchtman, N. (2012). Understanding peer effects in financial decisions: Evidence from a field experiment, *Econometrica, Forthcoming* .

Butler, J., Giuliano, P. and Guiso, L. (2010). The right amount of trust.

Butler, J. V., Giuliano, P. and Guiso, L. (2012). Trust and cheating, *Technical report*, National Bureau of Economic Research.

Calvó-Armengol, A., Patacchini, E. and Zenou, Y. (2009). Peer effects and social networks in education, *The Review of Economic Studies* **76**(4): 1239–1267.

Campbell, J. Y. and Cochrane, J. H. (1999). Vby force of habit: A consumption based explanation of aggregate stock market behaviorv, *Journal of Political Economy* **107**(2): 205–251.

Chandrasekhar, A. and Lewis, R. (2011a). Econometrics of sampled networks, *Unpublished manuscript, MIT.[422]* .

Chandrasekhar, A. and Lewis, R. (2011b). Econometrics of sampled networks. Working paper, MIT.

Chari, V. V. and Kehoe, P. J. (2004). Financial crises as herds: overturning the critiques, *Journal of Economic Theory* **119**(1): 128–150.

Chib, S. (1996). Calculating posterior distributions and modal estimates in markov mixture models, *Journal of Econometrics* **75**(1): 79–97.

Chib, S. and Greenberg, E. (1996). Markov chain monte carlo simulation methods in econometrics, *Econometric theory* **12**(03): 409–431.

Chipman, H. A., George, E. I., McCulloch, R. E. et al. (2010). Bart: Bayesian additive regression trees, *The Annals of Applied Statistics* **4**(1): 266–298.

Cliff, A. and Ord, J. (1973). *Spatial Autocorrelation*.

Cohen-Cole, E., Liu, X. and Zenou, Y. (2012). Multivariate choice and identification of social interactions, *CEPR Discussion Paper No. DP9159* .

Craig, B. (1999). Resisting electronic payment systems: burning down the house?, *Federal Reserve Bank of Cleveland* **4**.

Crump, R. K., Hotz, V. J., Imbens, G. W. and Mitnik, O. A. (2008). Nonparametric tests for treatment effect heterogeneity, *The Review of Economics and Statistics* **90**(3): 389–405.

Donald, S. and Newey, W. (2001). Choosing the number of instruments, *Econometrica* **69**(5): 1161–1191.

Drehmann, M., Goodhart, C. and Krueger, M. (2002). The challenges facing currency usage: will the traditional transaction medium be able to resist competition from the new technologies?, *Economic Policy* **17**(34): 193–228.

Edwards, F. (2012). Early to rise? the effect of daily start times on academic performance, *Economics of Education Review* **31**(6): 970–983.

Eide, E. R. and Showalter, M. H. (2012). Sleep and student achievement, *Eastern Economic Journal* **38**: 512–524.

Firpo, S. (2007). Efficient semiparametric estimation of quantile treatment effects, *Econometrica* **75**(1): pp. 259–276.

Fudenberg, D. and Maskin, E. (1986). The folk theorem in repeated games with discounting or with incomplete information, *Econometrica: Journal of the Econometric Society* pp. 533–554.

Gali, J. (1994). Keeping up with the joneses: Consumption externalities, portfolio choice, and asset prices, *Journal of Money, Credit and Banking* pp. 1–8.

Goldsmith-Pinkham, P. and Imbens, G. W. (2013). Social networks and the identification of peer effects, *Journal of Business and Economic Statistics* **31**: 253–264.

Granovetter, M. (1985). Economic action and social structure: the problem of embeddedness, *American journal of sociology* pp. 481–510.

Greif, A., Milgrom, P. and Weingast, B. R. (1994). Coordination, commitment, and enforcement: The case of the merchant guild, *Journal of political economy* pp. 745–776.

Gresvik, O. (2009). Costs in the payment system., *Economic Bulletin (Norges Bank)* **80**(1).

Guiso, L., Sapienza, P. and Zingales, L. (2004). The role of social capital in financial development, *American Economic Review* **94**(3): 526–556.

Guiso, L., Sapienza, P. and Zingales, L. (2008). Trusting the stock market, *The Journal of Finance* **63**(6): 2557–2600.

Hillman, D., Murphy, A., Antic, R. and Pezzullo, L. (2006). Economic cost of sleep disorders, *Sleep* **29**: 299–305.

Hong, H., Kubik, J. D. and Stein, J. C. (2004). Social interaction and stock-market participation, *The Journal of Finance* **59**(1): 137–163.

Hsieh, C. S. and Lee, L. F. (2011). A social interactions model with endogenous friendship formation and selectivity, *Technical report*, Working paper.

Hudgens, M. and Halloran, E. (2008). Toward causal inference with interference, *Journal of American Statistical Association* **103**: 832–842.

Humphrey, D. B. and Berger, A. N. (1990). Market failure and resource use: Economic incentives to use different payment instruments, *The US Payment System: Efficiency, Risk and the Role of the Federal Reserve*, Springer, pp. 45–86.

Humphrey, D. B., Pulley, L. B. and Vesala, J. M. (1996). Cash, paper, and electronic payments: a cross-country analysis, *Journal of Money, Credit and Banking* pp. 914–939.

Hung, A. A., Clancy, N., Dominitz, J., Talley, E. and Berrebi, C. (2008). *Investor and industry perspectives on investment advisers and broker-dealers*, Vol. 767, Rand Corporation.

Imai, K., Ratkovic, M. et al. (2013). Estimating treatment effect heterogeneity in randomized program evaluation, *The Annals of Applied Statistics* **7**(1): 443–470.

Imbens, G., Kolesar, M., Chetty, R., Friedman, J. and Glaeser, E. (2011). Identification and inference with many invalid instruments.

Imbens, G. and Woolridge, G. (2009). Recent developments in the econometrics of program evaluation., *Journal of Economic Literature* **47(1)**: 5–86.

Jackson, M. O., Rodriguez-Barraquer, T. and Tan, X. (2012). Social capital and social quilts: Network patterns of favor exchange, *The American Economic Review* **102**(5): 1857–1897.

Jackson, M. and Zenou, Y. (2014). *Economic Analyses of Social Networks*.

Kandori, M. (1992). Social norms and community enforcement, *The Review of Economic Studies* **59**(1): 63–80.

Karlan, D., Mobius, M., Rosenblat, T. and Szeidl, A. (2009). Trust and Social Collateral, *The Quarterly Journal of Economics* **124**(3): 1307–1361.

Kelejian, H. H. and Prucha, I. (2010). Spatial models with spatially lagged dependent variables and incomplete data, *Journal of Geographical Systems* **12**: 241–257.

Kelejian, H. and Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances, *The Journal of Real Estate Finance and Economics* **17**(1): 99–121.

Kelejian, H. and Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model, *International economic review* **40**(2): 509–533.

Kelejian, H. and Prucha, I. R. (2001). On the asymptotic distribution of the moran i test statistic with applications, *Journal of Econometrics* **104**(2): 219–257.

Kelejian, H. and Prucha, I. R. (2004). Estimation of simultaneous systems of spatially interrelated cross sectional equations, *Journal of Econometrics* **118**(1): 27–50.

Kelejian, H. and Prucha, I. R. (2007). Hac estimation in a spatial framework, *Journal of Econometrics* **140**(1): 131 – 154.

LeBlanc, M. and Kooperberg, C. (2010). Boosting predictions of treatment success, *Proceedings of the National Academy of Sciences* **107**(31): 13559–13560.

Lee, L. F. (2007). Identification and estimation of econometric models with group interactions, contextual factors and fixed effects, *Journal of Econometrics* **140**: 333–374.

Lee, L. F., Liu, X. and Lin, X. (2010). Specification and estimation of social interaction models with network structures, *The Econometrics Journal* **13**: 145–176.

Lin, X. (2010a). Identifying peer effects in student academic achievement by a spatial autoregressive model with group unobservables, *Journal of Labor Economics* **28**: 825–860.

Lin, X. (2010b). Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables, *Journal of Labor Economics* **28**(4): 825–860.

Lippert, S. and Spagnolo, G. (2011). Networks of relations and Word-of-Mouth Communication, *Games and Economic Behavior* **72**(1): 202–217.

Liu, X. (2013a). Estimation of a local-aggregate network model with sampled networks, *Economics Letters* **118**(1): 243–246.

Liu, X. and Lee, L. F. (2010). Gmm estimation of social interaction models with centrality, *Journal of Econometrics* **159**: 99–115.

Liu, X., Patacchini, E. and Rainone, E. (2013). The allocation of time in sleep: A social network model with sampled data, *CEPR Discussion Paper No. DP9752* .

Liu, X., Patacchini, E. and Zenou, Y. (2011). *Peer Effects in Education, Sport, and Screen Activities: Local Aggregate or Local Average?*, Centre for Economic Policy Research.

Liu, X., Patacchini, E. and Zenou, Y. (2014forthcoming). Endogenous peer effects: Local aggregate or local average?, *Journal of Economic Behavior and Organization* .

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem, *The review of economic studies* **60**(3): 531–542.

Manski, C. F. (2013). Identification of treatment response with social interactions, *The Econometrics Journal* **16**(1): S1–S23.

Miguel, E. and Kremer, M. (2004). Worms: Identifiying impacts on eduaction and health in the presence of treatment externalities, *Econometrica* **72**: 159–217.

Mitru, G., Millrood, D. L. and Mateika, J. H. (2002). The impact of sleep on learning and behavior in adolescents, *Teachers College Record* **104**: 704–726.

Patacchini, E. and Zenou, Y. (2008). The strength of weak ties in crime, *European Economic Review* **52**: 209–236.

Robert, C. P. and Casella, G. (2004). *Monte Carlo statistical methods*, Vol. 319, Citeseer.

Rubin, D. (1986). Which ifs have causal answers? discussion of hollands statistics and causal inference", *Journal of the American Statistical Association* **81**: 961–962.

Rubinstein, A. (1979). Equilibrium in supergames with the overtaking criterion, *Journal of Economic Theory* **21**(1): 1–9.

Sinclair, B., McConnell, M. and Green, D. P. (2012). Detecting spillover effects: Design and analysis of multilevel experiments, *American Journal of Political Science* **56**: 1055–1069.

Sojourner, A. (2013). Identification of peer effects with missing peer data: Evidence from project star, *The Economic Journal* **123**: 574–605.

Staiger, D. and Stock, J. H. (1997). Instrumental variables regression with weak instruments, *Econometrica* **65**(3): 557–586.

Tierney, L. (1994). Markov chains for exploring posterior distributions, *the Annals of Statistics* pp. 1701–1728.

Wang, W. and Lee, L. F. (2013a). Estimation of spatial autoregressive models with randomly missing data in the dependent variable, *Econometrics Journal* **16**: 73–102.

Wang, W. and Lee, L. F. (2013b). Estimation of spatial panel data models with randomly missing data in the dependent variable, *Regional Science and Urban Economics* **43**: 521–538.

Yaniv, G. (2004). Insomnia, biological clock, and the bedtime decision: an economic perspective, *Health Economics* **13**: 1–8.